Author:
Jefferies, Elizabeth Alice

Title:
Evidence for lexical and semantic contributions to phonological coherence in verbal short-term memory

Evidence for lexical and semantic contributions to phonological coherence in verbal short-term memory

Elizabeth Alice Jefferies

Department of Experimental Psychology

Submitted August 2003

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the aculty of Science

Word count: 76786

Abstract

This dissertation examines the contribution of long-term lexical and semantic representations to verbal short-term memory in both patients with semantic dementia, who have a specific and progressive loss of conceptual knowledge, and normal participants. Chapters 2, 3 and 4 compared the patients' recall of words that they comprehended relatively well and more poorly. The semantically degraded words were recalled less accurately and were characterised by more frequent phonological errors, consistent with the view that semantics plays a major role in maintaining phonological integrity in normal immediate recall. However, several previous studies have failed to obtain a known-degraded recall difference, challenging this view. Chapter 2 examined the effect of various methodological factors on the magnitude of the known-degraded recall difference and found that set size could potentially explain much of this discrepancy in results. Chapter 4 examined the evidence for phonological-lexical deficits independent of the patients' primary semantic impairment and concluded that substantial known-degraded differences can emerge even in the absence of phonological impairment. Chapter 3 considered the patients' immediate recall of number and non-number words; number words were comparatively free from phonological errors and were comprehended relatively well, suggesting that the patients' category specific advantage for numbers in verbal short-term memory may have had a semantic locus. Chapter 5 examined the impact of lexical and semantic factors on the immediate recall of healthy participants; they made phonological errors like those displayed by SD patients when they recalled mixed lists of words and nonwords. In Chapter 6, both patients and normal participants showed an effect of lexical/semantic variables in recognition as well as recall. These results are discussed in terms of interactive and late-stage redintegration accounts of the long-term memory contribution to verbal short-term memory. It is argued that several findings across the chapters are more consistent with the interactive viewpoint.

Acknowledgements

Professor Matthew Lambon Ralph, Professor Alan Baddeley and Dr. Clive Frankish supervised the work in this dissertation. I would like to thank them for their unfailing support of my research and for their generosity with their time and energy. I would also like to thank Dr. Karalyn Patterson for many interesting discussions about issues raised within this thesis and for suggesting the use of the mixed words-nonwords methodology adopted in Chapter 5. Thanks are due to Manabu Ikeda for sharing the results from the mathematical calculation assessments presented in Chapter 3 and to Marinella Cappelletti and Brian Butterworth for allowing me to make use of their number word reading materials. Computer software designed by Clive Frankish was used to construct the nonwords used in Chapters 5 and 6 and to look up the properties of many of the experimental items in the Celex database. I would especially like to thank EK, GT, PD, MK, SJ, KI, JT and BS for their kind participation in this research. I am also indebted to Dr. Roy Jones, St Martin's Hospital, Bath and Dr. David Bateman, Royal United Hospital, Bath for allowing us access to these patients. I gratefully acknowledge that this research was supported by an ESRC studentship.

Authors' Declaration

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original except where indicated by special reference in the text.

Any views expressed in the dissertation are those of the author and in no way represent those of the University of Bristol.

The dissertation has not been presented to any other University for examination either in the United Kingdom or overseas.

Signed: Elizabeth Jeffers

Date: 31/1/04

Contents

Chapter 2: When does word meaning affect immediate serial recall in semantic dementia?

Chapter 3: A category-specific advantage for numbers in verbal short-term memory

Chapter 4: Evidence for intact phonology in semantic dementia

Chapter 5: Lexical and semantic factors impact on phonological coherence in normal immediate serial recall

Chapter 6: Lexical and semantic influences on immediate serial recognition

Chapter 7: Discussion and conclusions

List of tables and illustrations

Chapter 2: When does word meaning affect immediate serial recall in semantic dementia?

Chapter 3: A category-specific advantage for numbers in verbal short-term memory

Chapter 4: Evidence for intact phonology in semantic dementia

Chapter 5: Lexical and semantic factors impact on phonological coherence in normal immediate serial recall

Chapter 6: Lexical and semantic influences on immediate serial recognition

1 Verbal short-term memory and stable linguistic representations: A review of the literature

1.1 Introduction

Models of verbal short-term memory (STM) have traditionally placed particular emphasis on the role of phonological representations in immediate serial recall (ISR). It is now commonly acknowledged, however, that verbal STM draws on multiple levels of representations (e.g., lexical, semantic and syntactic) that play a role in language production and comprehension. Despite this consensus, the exact nature of this contribution remains controversial and is the primary focus of this thesis. This chapter reviews the experimental and neuropsychological evidence for a role of linguistic representations beyond phonology in ISR performance. Initially, the focus is on the role of lexical-level representations in the recall of lists of unrelated words, but later, the role played by super-lexical processes in the immediate recall of sentences is considered. This is followed by a discussion of the nature of the relationship between stable linguistic representations and verbal STM, in which different theoretical approaches are described and contrasted.

1.2 Multiple levels of stable linguistic representations contribute to verbal short-term memory

1.2.1 Evidence for a short-term phonological code

It is well established that verbal STM draws heavily on a phonological code, accounting for the occurrence of phonological errors in ISR tasks (Conrad, 1964) and the poorer recall of phonologically similar than dissimilar items (Conrad & Hull, 1964). In line with these findings, Baddeley's highly influential Working Memory (WM) model (Baddeley, 1986; Baddeley & Hitch, 1974) proposes that verbal STM is underpinned by a phonological loop subsystem, operating independently of LTM. The phonological loop comprises two components: a phonological store that represents information in a rapidly decaying phonological code, and an articulatory loop that uses subvocal rehearsal to refresh the contents of the store. Articulatory suppression, in which participants utter irrelevant material like "the, the, the", greatly reduces ISR for printed words (Estes, 1973; Levy, 1971; Murray, 1967; Peterson & Johnson, 1971), purportedly because it prevents subvocal rehearsal. In addition, ISR is better for shorter items that can be articulated more quickly (Baddeley, Thomson, & Buchanan, 1975; Ellis & Hennelly, 1980), supporting the notion that the phonological loop has a time-based capacity linked to an articulatory rehearsal process.

Although the phonological loop model is based on a large body of empirical data (see Gathercole & Baddeley, 1993, for a review), it does not offer a satisfactorily account of the way in which stable linguistic representations contribute to verbal STM. Span for verbal materials increases with the size of the contribution from LTM, and can rise considerably above the supposed capacity of the phonological loop. The difference between word and sentence span provides a particularly clear example of this; twelve or more words can be recalled in the correct order when they form a meaningful sentence, whereas span for unrelated words is limited to around half that number (Brener, 1940).

1.2.2 The role of stable lexical-level representations in short-term memory

1.2.2.1 Evidence from healthy participants

1.2.2.1.1 Effects of lexicality and frequency

Different levels of long-term knowledge (sub-lexical, lexical, super-lexical) appear to contribute to performance on ISR tasks. The role of stable lexical-level representations in verbal STM is clearly demonstrated by studies that have found substantially better ISR for words than for nonwords (Brener, 1940; Hulme, Maughan, & Brown, 1991; Hulme, Roodenrys, Brown, & Mercer, 1995; Roodenrys, Hulme, & Brown, 1993; Turner, Henry, & Smith, 2000). Moreover, Collette et al. (2001) showed that verbal STM for words recruits additional cortical areas associated with lexical and semantic processing (left middle temporal gyrus and temporo-parietal junction) compared with verbal STM for nonwords, in a study using positron emission tomography (PET).

Although the lexicality effect in ISR could be underpinned by either long-term phonological or semantic representations, independent effects of both types of coding can be demonstrated. Hulme et al. (1995) found that familiarising participants with the phonological forms of nonwords improved their recall, even though these items remained meaningless. In addition, ISR is higher for words that occur frequently in English compared with less frequent words (Gregg, Freedman, & Smith, 1989; Hulme et al., 1997; Roodenrys, Hulme, Alban, Ellis, & Brown, 1994; Roodenrys & Quinlan, 2000; Tehan & Humphreys, 1988; M. J. Watkins, 1977; O. C. Watkins & Watkins, 1977). Although articulation rates are faster for high frequency words (Wright, 1979), Tehan and Humphreys (1988) and Gregg et al. (1989) showed that the superior recall of high frequency words persists under conditions of articulatory suppression. In addition, Hulme and colleagues demonstrated effects of both of lexicality and frequency on ISR that were at least partially independent of speech rate (Hulme et al., 1991; Hulme et al., 1997). This research, therefore, provides an example of variation in ISR across materials that cannot be adequately accounted for by differences in articulatory rehearsal as predicted by Baddeley's phonological loop model. Frequency effects in ISR are considered to be lexical-phonological in nature; the stable phonological representations of high frequency words are thought to be more accessible or better specified than their low frequency counterparts (Hulme et al., 1997). Both lexicality and frequency appear to selectively affect the rate of item identity errors in ISR. In contrast, these variables have little impact on the occurrence of order errors, in which items are recalled in the wrong serial positions (Gathercole, Pickering, Hall, & Peaker, 2001; Hulme et al., 1997).

1.2.2.1.2 Semantic effects

Semantic representations also contribute to ISR for word lists. Walker and Hulme (1999) found that highly concrete words were recalled more accurately than more abstract words. This result, which was not attributable to differences in speech rate, affected the frequency of item identity but not order errors. Similarly, Bourassa and Besner (1994) found superior ISR for content compared with function words. This effect persisted under conditions of articulatory suppression suggesting that it did not result from variation in articulation rates. The difference was entirely eliminated when the content and function words were matched for imageability, however, suggesting that this factor, and not the difference in grammatical class, underpinned the effect. Highly imageable/concrete words have been assumed by a number of authors to have 'richer' semantic representations than less imageable words, perhaps because they are associated with a larger number of semantic features (Jones, 1985; Plaut & Shallice, 1991). These findings therefore suggest that semantic representations play a part in verbal STM.

Although STM tasks are generally found to be less sensitive to semantic than phonological similarity (Baddeley, 1966, 1972; Shulman, 1971), a number of studies have observed superior ISR for words drawn from one rather than several semantic categories, consistent with a semantic contribution to ISR (Brooks & Watkins, 1990; Huttenlocher & Newcombe, 1976; Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier, 1999; Wetherick, 1975). In contrast with these findings, Baddeley (1966) found slightly *poorer* immediate recall for semantically similar than dissimilar items. This apparent contradiction may have arisen because the facilitatory effect of semantic similarity is largely restricted to item recall (i.e., has little impact on the occurrence of order errors: Poirier & Saint-Aubin, 1995). Baddeley's task, on the other hand, primarily tapped order memory because the set size was highly restricted allowing participants to become familiar with the entire set of items. There is still debate about whether semantic factors can impinge on STM for serial order in lists of unconnected words (see Poirier and Saint Aubin, 1999). However, Baddeley and Ecob (1970) found that order recall was considerably better for words that formed a meaningful phrase (e.g., 'I might fly' or 'might I fly'), compared with semantically unrelated words (e.g., 'eye, fight, dry').

suggesting that stable semantic/syntactic representations do support the retention of order in word lists that approximate sentences.

1.2.2.2 Neuropsychological evidence

Neuropsychological evidence also points to the role of semantic representations in span tasks. In several studies, Martin and colleagues (R. C. Martin & Lesch, 1996; R. C. Martin, Lesch, & Bartha, 1999; R. C. Martin, Shelton, & Yaffee, 1994) have described patients with an apparently specific STM impairment for phonological or semantic information resulting from cardiovascular accident (CVA) or herpes encephalitis. For example, R. C. Martin et al. (1994) compared the performance of two patients, AB and EA, who appeared to have particular impairments in their retention of semantic and phonological information respectively. AB performed poorly on a category probe task requiring him to judge if a probe word was in the same category as any of the items in a preceding list, but much more accurately at a rhyme probe task. In contrast, EA performed well on the category probe task but more poorly on the rhyme probe task. In line with this, AB showed a normal effect of phonological similarity in his ISR performance but a reduced lexicality effect, whereas EA showed the opposite pattern. AB's results were replicated in a second patient, MS, with a similar semantic retention deficit (R. C. Martin et al., 1999). MS was additionally found to show enhanced effects of frequency and imageability in ISR. N. Martin and Saffran (1997) obtained comparable results in a group study that examined 15 aphasic patients with semantic or phonological processing impairments. The patients' semantic abilities, but not their phonological abilities, were related to the magnitude of frequency and imageability effects in ISR. These results suggest that phonological and semantic codes are both involved in verbal STM and can be impaired separately.

Even more compelling evidence for a lexical-semantic contribution to verbal STM is provided by studies that demonstrate that, for an individual patient, the likelihood of recalling a particular word in an ISR task is affected by the degree to which it is semantically degraded, according to performance on semantic tests like naming and word-picture matching. This methodology has been used most frequently with patients with semantic dementia (SD), a neurodegenerative condition that produces a specific decline in semantic memory. SD is the temporal variant of frontal-temporal dementia, and is associated with progressive focal atrophy of the infero-temporal neocortex which is typically more pronounced in the left hemisphere (Hodges, Patterson, Oxbury, & Funnell, 1992; Snowden, Goulding, & Neary, 1989). SD patients are anomic and have impaired comprehension on both verbal and non-verbal tasks. However, their perceptual and spatial skills, new episodic learning, non-verbal reasoning, syntax and phonology remain largely intact (Hodges et al., 1992). SD patients almost never produce phonological errors in spontaneous speech, have intact digit span and generally perform well on phonological tasks like minimal pair discrimination (Knott, Patterson, & Hodges, 1997). However, ISR is characterised by phonological breakdown in which phonemes migrate to new positions in the list and word span is severely impaired (Patterson, Graham, & Hodges, 1994).

Several studies have found that this phonological breakdown is markedly more severe for words that are understood poorly compared with those that are better known supporting the view that semantic representations make a substantial contribution to verbal STM. Patterson et al. (1994) found considerable recall differences between relatively well-known and semantically degraded words for three patients with SD. Although the known and degraded words used in this study were not matched for word frequency, the differences remained significant when the highest frequency known items were discarded. Several subsequent studies have found a substantial recall difference between known and degraded words matched for frequency on an item-by-item basis (Knott et al., 1997; Knott, Patterson, & Hodges, 2000). In addition to these studies of SD patients, ISR differences between relatively well-known and semantically degraded words have been reported in semantically impaired patients following cardiovascular accident (Forde & Humphreys, 2002) and herpes simplex encephalitis (Caza, Belleville, & Gilbert, 2002; R. C. Martin et al., 1999). However, there have also been some notable failures to find such differences (Funnell, 1996; Lambon Ralph & Howard, 2000; McCarthy & Warrington, 1987, 2001; Warrington, 1975) and possible reasons for these discrepancies in results are discussed in Chapter 2.

1.2.3 Sub-lexical effects on short-term memory

The literature reviewed above provides convincing evidence to suggest that stable phonological and semantic representations play a substantial part in verbal STM. In addition to these lexical-level effects however, LTM may make a contribution to ISR at a sub-lexical level. Gathercole, Frankish, Pickering and Peaker (1999) found an effect of phonotactic frequency on the recall of nonwords; nonwords containing common phoneme combinations (e.g., "riss") were recalled more accurately than those containing more unusual combinations of phonemes (e.g., "youdge"). Several authors have argued that phonotactic frequency makes an important contribution to 'wordlikeness', that is, the degree to which a nonword is judged to resemble a real word (Bailey & Hahn, 2001; Gathercole & Martin, 1996). As wordlikeness has a substantial impact on the recall of nonwords (Gathercole, 1995; Gathercole, Willis, Emslie, & Baddeley, 1991; Van Bon & Van der Pijl, 1997), these findings support the notion of a sub-lexical effect in nonword recall. An alternative explanation of this relationship, however, is that more word-like nonwords have a larger number of phonologically similar real-word neighbours (e.g., "cat", "fat" and "hat" for "lat"), and therefore the wordlikeness effect may be lexically mediated. Gathercole et al.'s (1999) materials were not controlled for neighbourhood size. Roodenrys and Hinton (2002) found that the phonotactic frequency effect disappeared when neighbourhood size was held constant. In contrast, an effect of neighbourhood size remained even when phonotactic frequency was controlled. This study suggests that lexical representations may play a role in nonword recall; a possibility that is discussed at greater length in Chapter 4.

1.2.4 Super-lexical effects in short-term memory

As mentioned previously, sentence span is substantially greater than word span (Brener, 1940), reflecting the involvement of long-term representations above the lexical level in span tasks. Miller and Selfridge (1950) studied the recall of word sequences with different degrees of approximation to English prose. Higher order approximations were more meaningful and more grammatical, and resulted in better recall, suggesting that both syntactic and semantic representations may play a role in increasing sentence span above word span. Evidence for the involvement of syntactic factors is provided by a study by Epstein (1961). Grammatically marked sequences of nonsense syllables and unrelated words were easier to recall than sequences that were not tagged in this way (e.g., "the yigs wur vumly rixing" was easier to learn than "the yig wur vum rix"). Similarly, Marks and Miller (1964) found that 'anomalous sentences' that were grammatically correct but had little overall meaning (e.g., "Noisy flashes emit careful floods") were recalled more accurately than jumbled anomalous sentences with a random word order.

The semantic relationships between words can also influence STM. Baddeley and Levy (1971) found a semantic similarity effect in the recall of meaningful noun-adjective pairs, so that sequences like "priest-moral, minister-religious, vicar-pious" were recalled more poorly than sequences like "palace-magnificent, apple-delicious, rattlesnake-deadly". This semantic similarity effect disappeared when the nouns and adjectives were not semantically compatible (e.g., "palace-moral, castle-religious, fort-pious"). In addition, information about the overall theme of a passage of prose in the form of a title or a picture improves recall considerably (Bransford & Johnson, 1972; Dooling & Lachman, 1971), providing further evidence for a role of conceptual or thematic knowledge in the immediate recall of prose.

More recently, Potter and Lombardi (Lombardi & Potter, 1992; Potter & Lombardi, 1990, 1998) obtained convincing evidence for a role of semantic and syntactic representations in the immediate recall of sentences, in a series of studies looking at false recall and recognition. Participants read sentences like "The knight rode around the palace searching for a place to enter", followed by a list of unrelated words, one of which was semantically related to a word in the sentence (e.g., "castle"). When participants attempted to recall the sentence, they frequently falsely produced the 'lure' word (castle) in the place of its synonym (palace). They also readily accepted the lure word as having been in the sentence in a recognition paradigm (Potter & Lombardi, 1990). Similar experiments suggested that syntactic representations were also involved in the immediate recall of sentences (Lombardi & Potter, 1992; Potter & Lombardi, 1998). For example, participants presented with two clause sentences often incorrectly recalled the first clause following the surface syntactic structure of the second clause (Potter & Lombardi, 1998).

Potter and Lombardi argue from these findings that the surface structure of sentences can be reconstructed from the long-term conceptual, lexical and syntactic representations activated during sentence processing. They advocate an extreme view in which phonological activation does not contribute to short-term sentence recall. This claim is consistent with studies showing that errors in sentence recall often preserve gist and not surface structure (e.g., Jarvella, 1971). However, this pattern is delay dependent; memory for surface structure deteriorates rapidly with a filled delay, whereas gist memory is relatively unaffected (Sachs, 1967). In addition, Potter and Lombardi's claim is apparently incompatible with neuropsychological evidence demonstrating substantial sentence repetition deficits in patients with phonological STM impairments (e.g., Butterworth, Campbell, & Howard, 1986; Hanten & Martin, 2000; R. C. Martin, 1993; R. C. Martin et al., 1994; Saffran & Martin, 1990) and has been called into question by recent studies demonstrating phonological effects in sentence repetition in intact individuals (Rummer & Engelkamp, 2001, 2003; Willis & Gathercole, 2001). An alternative view is that the immediate verbatim repetition of sentences draws on all the types of representations that are involved in sentence processing – phonological as well as semantic and syntactic.

1.3 Theoretical perspectives on the nature of the long-term memory contribution to short-term memory

The evidence reviewed thus far clearly demonstrates that long-term phonological, semantic and syntactic representations play a substantial role in verbal STM. Although there is now widespread agreement that long-term representations do contribute to verbal STM, the nature of this contribution is still highly controversial.

Theories about the relationship between LTM and STM are numerous and wide-ranging. Broadly speaking, however, the different perspectives can be grouped together into three categories. First, some theorists view verbal STM as temporary activation of linguistic representations in LTM (e.g., MacDonald & Christiansen, 2002; N. Martin & Saffran, 1997; Patterson et al., 1994). Verbal STM is not seen as an autonomous cognitive system but rather as ongoing activity within the language system itself. According to this perspective, patients with linguistic processing deficits should show closely related verbal STM impairments. A second perspective retains the close relationship between the language system and verbal STM while postulating separate short-term and long-term stores (e.g., R. C. Martin & Lesch, 1996; R. C. Martin et al., 1999). Word span tasks activate stable lexical representations but the words are also temporarily encoded in an independent STM system. This account predicts the existence of patients with impaired verbal STM but normal linguistic processing. A third perspective advocates an even greater divergence between verbal STM and long-term linguistic representations. Verbal STM is seen as independent of the language system although there are interactions between transient representations in STM and stable representations in LTM that allow linguistic knowledge to enhance verbal STM performance (Baddeley, Gathercole, & Papagno, 1998). In particular, several theorists have suggested that when the transient phonological trace becomes degraded, it can be 'cleaned-up' using LTM (Hulme et al., 1991; Hulme et al., 1997; Schweickert, 1993; Schweickert, Chen, & Poirier, 1999). In this section, each of these perspectives will be discussed in turn.

It should also be noted that some theorists, most notably Monsell (1984), have proposed that there are several types of STM, corresponding to these different mechanisms. Monsell characterised STM in terms of two types of temporary storage: type I is the persisting activation of stable representations and type II allows the representation of novel structure and may involve a separate STM buffer. The type I mechanism was proposed to be more involved in the retention of item information in ISR tasks, and the type II mechanism was suggested to play a greater role in memory for serial order.

1.3.1 Verbal short-term memory as temporary activation of stable linguistic representations

Verbal STM may correspond to the temporary activation of stable linguistic representations, rather than a separate cognitive system (MacDonald & Christiansen, 2002; N. Martin & Saffran, 1997; Patterson et al., 1994). This approach has its roots in the general view that the temporary storage of information emerges from the systems that process that information (Cantor & Engle, 1993; Craik & Lockhart, 1972; Crowder, 1993; Hebb, 1949; McClelland & Rumelhart, 1985). It seems likely that the language system has some short-term storage capacities, as brief storage appears to be necessary for comprehending and producing utterances. This perspective naturally accounts for the role played by long-term phonological, semantic and syntactic representations in verbal STM, as activation within each of these types of linguistic representation will contribute to the overall performance of the system. The different representational levels are thought to interact so that activation at one level can mutually constrain activity at the other levels. This property helps the system to settle on accurate, self-reinforcing patterns of activation.

1.3.1.1 The interactive-activation account (N. Martin and Saffran)

N. Martin and Saffran (1997) have adopted an approach of this nature. They sought to account for the verbal STM performance of patients with linguistic impairments with reference to Dell and O'Seaghda's (1992) interactive activation (IA) model of word production. This model assumes that temporary storage is an inherent property of the language processor, as activation is maintained across a number of processing cycles until a response is produced. Maintenance of the order of several words is, however, beyond the scope of the model. The model contains three distinct processing levels: phonological, lexical and semantic. Each level consists of localist nodes, linked by bi-directional connections to the nodes in adjacent levels. Activation spreads forwards and backwards between the levels during every processing cycle. In word production, semantic activation feeds down to the lexical and phonological nodes, whereas in comprehension, phonological activation spreads up to the semantic nodes. In repetition, phonological activation spreads up to the lexical and semantic levels and then down again. Consequently, the lexical and semantic levels contribute to verbal STM by helping to sustain activation in the phonological nodes, which is prone to rapid decay.

N. Martin and colleagues used this IA approach to account for the naming, repetition and verbal STM impairments of an aphasic patient, NC (N. Martin, Dell, Saffran, & Schwartz, 1994; N. Martin & Saffran, 1992; N. Martin, Saffran, & Dell, 1996). When he was first tested, NC made a predominance of formal paraphasias in naming. His single word repetition was dominated by semantic errors and he was unable to repeat nonwords. In addition, his verbal STM span was restricted to a single item. Martin and Saffran (N. Martin & Saffran, 1992) were able to account for this apparently diverse pattern of impairments within Dell's framework described above, by positing a pathological increase in the rate of decay of activation across all the nodes in the network. In the model, rapid decay increases the probability that phonologically or semantically related items are incorrectly selected for output, accounting for NC's unusual pattern of errors. As NC recovered, he made fewer semantic errors in repetition, and a larger number of formal paraphasias and neologisms. Simulations showed that the IA model mirrored this pattern of recovery as the decay rate was decreased (N. Martin et al., 1994). NC's error pattern returned to its pre-recovery state when repetition was delayed, consistent with simulations showing that decay effects within the model are exaggerated by delays (N. Martin et al., 1996). These studies show that errors in naming, repetition and verbal STM can be accounted for parsimoniously by assuming that verbal STM is an emergent property of the language system.

1.3.1.2 The semantic binding hypothesis (Patterson et al.)

Patterson and colleagues (Knott et al., 1997; Patterson et al., 1994) have adopted a complimentary approach. Their 'semantic binding hypothesis' also suggests that verbal STM emerges from interactions between different types of linguistic representation. In contrast to the model of N. Martin, however, this hypothesis has its roots in the parallel distributed processing (PDP) models of, for example, Seidenberg and McClelland (1989), which posit distinct semantic and phonological representations but not a separate lexical level. According to the semantic binding hypothesis, there are two sources of coherence that play a role in producing the correct configuration of phonological elements in both speech production and verbal STM. Firstly, because the elements of a word are always activated together when that word is produced, they become associated in the phonological system. Consequently, the phonological system develops pattern completion properties for familiar words. A second source of coherence is provided by the semantic system. Every time a word is spoken or comprehended, semantic activation co-occurs with activation representing the phoneme sequence for that word. As a result, semantics can constrain the pattern of activation in the phonological system, and increase the likelihood that the phonemes of words are produced in the correct order.

1.3.1.3 Evidence for interactive models

The models of N. Martin and Patterson predict that there should be a close association between linguistic impairments and deficits in verbal STM, as verbal STM is seen as relying on the representations that underlie language processing. Several studies were reviewed earlier that examined the verbal STM of patients with either phonological or semantic deficits (N. Martin & Saffran, 1997; R. C. Martin & Lesch, 1996; R. C. Martin et al., 1999; R. C. Martin et al., 1994). The phonologically impaired patients showed a greater effect of semantic factors in ISR, whereas the semantically impaired patients showed a greater effect of phonological factors in ISR, in line with this prediction. Moreover, several studies have found that semantically impaired patients have superior ISR for words that they understand relatively well, compared with words that are more semantically impaired (Caza et al., 2002; Forde & Humphreys, 2002; Knott et al., 1997, 2000; Patterson et al., 1994), again suggesting a close relationship between the status of linguistic representations and verbal STM performance. However, as described below, models that postulate separate STM and LTM stores can also accommodate these findings.

As the semantic binding hypothesis specifically proposes that semantic activation helps to bind the phonological elements of words together, a lack of semantic binding should produce a pattern of errors in which the phonemes of semantically degraded words migrate between list items. Several studies have shown that patients with semantic impairments do make a greater number of phoneme migration errors in their recall of words that they understand poorly (Caza et al., 2002; Forde & Humphreys, 2002; Knott et al., 1997, 2000; Patterson et al., 1994). Interestingly, normal participants show a similar pattern of phoneme order errors in their serial recall of non-words that by definition lack lexical and semantic representations (Treiman & Danis, 1988). Phoneme migrations can also occur in the normal recall of word lists, particularly when the words are not repeated in the course of the experiment (Gathercole et al., 2001).

If, as suggested by the semantic binding hypothesis, semantics helps to maintain the phonology of words in ISR, it should play a similar role in other apparently 'non-semantic' tasks requiring phonological production, for example, reading aloud. Some views about the translation from orthography to phonology suggest that semantic representations play an important role in reading aloud, especially for low frequency words with atypical spelling-to-sound correspondences (Plaut, McClelland, Seidenberg, & Patterson, 1996). SD patients make reading errors on such words, pronouncing them as if they had regular correspondences (PINT to rhyme with "mint"): i.e., they demonstrate surface dyslexia (Graham, Hodges, & Patterson, 1994; Patterson & Hodges, 1992). In line with these findings, normal participants show effects of imageability on their reading times for single low frequency irregular words (Strain, Patterson, & Seidenberg, 1995). The approach adopted by N. Martin and Patterson also predicts that semantics should play a role in single word repetition, as single word repetition and ISR tasks are seen as being underpinned by the same language system. Such effects have been demonstrated:

for example, Tyler, Voice and Moss (1996) found an effect of imageability on the latency of single word repetition in healthy participants. This result parallels the effects of imageability in verbal STM.

1.3.2 A close relationship between language processing and verbal short-term memory but separate short-term and long-term stores

R. Martin and colleagues have advocated a rather different view of the relationship between STM and LTM (R. C. Martin & Breedin, 1992; R. C. Martin & Lesch, 1996; R. C. Martin et al., 1999; R. C. Martin et al., 1994). Their 'multiple components' theory concurs with the viewpoint discussed above in proposing a close association between the representations involved in language processing and verbal STM. Consequently, phonological, lexical and semantic representations are expected to be involved in both language processing and ISR. In contrast to the models discussed above, however, this approach proposes that separate systems underlie short-term storage and long-term knowledge. Just as several types of stable linguistic representation play a role in language processing, R. Martin proposes that there are several types of short-term buffer (i.e., phonological, semantic) that contribute to verbal STM. The short-term buffers are thought to have strong links with their corresponding long-term linguistic representations. Inputs initially activate these stable linguistic representations, which are conceived of in terms of the IA model of Dell and O'Seaghda (1992), but are then also temporarily encoded in the short-term buffers.

Although it could be argued that this view is insufficiently parsimonious, R. Martin maintains that a distinction between LTM and STM is required to account for certain neuropsychological dissociations. This account, unlike those discussed previously, predicts that it is possible for verbal STM to be impaired in the context of intact language processing. In line with this prediction, R. Martin et al. (R. C. Martin & Lesch, 1996; R. C. Martin et al., 1994) claimed that their patient AB had a specific deficit in the retention of semantic information, despite having apparently intact semantic processing. AB was impaired at a category probe test, which required the retention of semantic information, despite performing well on tests on naming and word-picture matching. In addition, he performed poorly on attribute questions like "Which is soft, cotton or sandpaper?" but only when they were presented auditorily. When the same questions were presented visually, AB performed perfectly, suggesting his knowledge of the attributes of objects was intact but his ability to retain semantic information was impaired. One problem with this interpretation, however, is that tasks that require the retention of semantic information may be more demanding than those that do not, enabling them to reveal subtle semantic processing deficits that would otherwise pass unnoticed. This issue is discussed at greater length in section 1.4.1 below.

This suggestion of multiple buffers for different types of linguistic representation has similarities to the model presented by Barnard (1985). In Barnard's production system model, when an input activates a particular 'production' (for example, when an auditory input activates a lexical form), a parallel copy of the input is made in an 'image record' that is specialized for that type of input. Information in the image record can be re-presented to the production system if off-line interpretation is required.

McCarthy and Warrington (1987) have also proposed that separate STM systems operate for different types of input. According to these authors, word span is underpinned by a phonological STM store that is sensitive to lexicality but independent of semantic knowledge (also see McCarthy & Warrington, 2001). In contrast, sentence repetition is supported by a dynamic, integrative memory system that draws heavily on semantics. McCarthy and Warrington pointed to a double dissociation between two verbal STM impaired cases and an SD patient as support for this theory. The STM patients were markedly impaired on word span tasks but not sentence repetition, whereas the SD patient was relatively unimpaired on short word lists but more strikingly impaired on sentence repetition, particularly when the sentences contained words he did not understand. The SD patient did not show a difference between relatively well understood and more semantically degraded words in his word list recall performance, suggesting to McCarthy and Warrington that the phonological STM store is not influenced by semantic factors. However, several other studies have reported known-degraded differences in ISR for word lists (see section 1.2.2.2).

1.3.3 Verbal STM is independent of the language system

A third approach, most closely aligned with the WM tradition, views verbal STM as an autonomous cognitive system (e.g., Baddeley, 1986; Baddeley et al., 1998; Shallice, 1988; Shallice & Warrington, 1970). The transient phonological code of the verbal STM system is thought to be distinct from the stable phonological representations of the language system and underpinned by different brain structures. Although stable linguistic representations are still purported to contribute to verbal STM, their role is a much more minor one; for example, several authors have suggested that LTM only facilitates STM when the phonological trace has become degraded (Hulme et al., 1991; Hulme et al., 1997; Schweickert, 1993; Schweickert et al., 1999).

Baddeley and Gathercole (1998) proposed that there are separate short-term and long-term phonological representations in order to account for the role of verbal STM in vocabulary acquisition (e.g., Baddeley, Gathercole & Papagno, 1998; Gathercole & Baddeley, 1990; Papagno, Valentine, & Baddeley, 1991) and the impact of lexical knowledge on verbal STM (e.g., Hulme et al., 1991). They suggested that both familiar and unfamiliar verbal inputs are stored temporarily in a short-term store that has the capacity to modify weights in the long-term store, allowing new sound sequences to be learned. Representations in the short-term store degrade rapidly but can be reinstated with reference to the long-term representations, in a process of 'redintegration'. Long-term representations are therefore seen as playing a non-essential and late, reconstructive role in verbal STM. Schweickert (Schweickert, 1993; Schweickert et al., 1999) and Hulme et al. (1997) have also suggested that LTM contributes to STM through a process of redintegration. According to this viewpoint, intact phonological traces can be retrieved directly from the short-term store, without reference to LTM, but more degraded traces are reinstated through a process of redintegration just prior to overt recall. This theory seems to imply that the redintegration mechanism can be strategically turned off for nonwords – otherwise these items would be erroneously subjected to the reconstruction process. However, it is not clear how the system could distinguish the degraded phonological trace of a word from that of a nonword. The approaches of Patterson et al. and Martin and Saffran, in contrast, do not require any additional assumptions. According to these theories, the initial processing of words produces additional lexical and semantic activation that is not elicited by nonwords, and this activation supports ISR for words over nonwords.

Models that view verbal STM as an emergent property of the language system predict that the full range of linguistic representations, phonological, semantic and syntactic, will contribute to verbal STM. In contrast, redintegration theories do not lend themselves to a straightforward account of the effect of semantic representations on verbal STM, although they can accommodate them with certain modifications. Walker and Hulme (1999) suggested that a semantic redintegration effect could operate in parallel with phonological redintegration. Long-term semantic representations could be used to reinstate short-term semantic activation in much the same way that long-term phonological representations are thought to reinstate the phonological trace (this suggestion has similarities with the notion of multiple buffers favoured by R. Martin and colleagues). Poirier and Saint Aubin (1995) alternatively proposed that semantic activation could help to constrain the phonological representations that were seen as candidates in the reconstruction process. For example, in a list in which all the items are animals, "_at" is likely to be reconstructed as "cat" and not "hat".

1.3.4 Connectionist models of phonological STM

Three distinct viewpoints of the relationship between verbal STM and long-term linguistic representations have now been described. Before comparing them in more detail, some consideration should be given to connectionist models of phonological STM and how they relate to these different perspectives.

Burgess and Hitch (1992; 1996; 1999) have proposed a connectionist model of the phonological loop component of WM that is a hybrid of the viewpoints discussed above. The model consists of three sets of localist units representing phonemes, items and a time varying context signal. The context signal becomes associated with item activation through Hebbian learning, allowing ordered recall to occur when the signal is replayed. When a verbal stimulus is presented, the phoneme units become active, and these in turn activate the item units. During recall, the item units are re-activated by the context signal and then compete for selection. The single winner of this competition reinstates the appropriate phoneme activation for that item, and consequently, either the correct item is recalled or an item order error occurs.

The original version of this model (Burgess & Hitch, 1992) was able to reproduce many critical properties of serial recall, such as the bow-shaped serial position curve and frequent item order errors between neighbouring items, but it could only handle familiar words. The model did not include a mechanism for learning over repeated trials and consequently it did not provide a satisfactory account of the effects of prior experience on ISR. The 1999 model, in contrast, includes 'fast' and 'slow' learning weights in the connections between 1) items and context units and 2) items and phonemes (similar to those used by Plaut & Shallice, 1993). The decay of fast weights accounts for loss of information from STM, whereas the slow weights explain the effects of item familiarity (e.g., the lexicality effect) and list repetition (the Hebb effect). Word recall is superior to nonword recall because stronger slow connections exist between familiar items and their consistent phonemes. It is proposed, therefore, that lexical-level item representations restore the appropriate phonological activation for words during the process of recall, in a manner that resembles redintegration. However, the model also suggests that the same units underpin STM and LTM, in line with the first perspective discussed above.

Glasspool's (1995) model functions in a similar way to the Burgess and Hitch (1992) model but has parallel mechanisms for the ordering of items and phonemes in ISR. The inclusion of a context pattern and competitive filter for the phoneme nodes allows the model to account for phoneme order errors, such as those that occur commonly in normal nonword recall (Treiman & Danis, 1988) and the ISR of semantically degraded words in SD patients (Patterson et al., 1994). Verbal inputs are presented simultaneously to the phoneme and word nodes in this model, but the item nodes will only become active if the input includes recognisable words. During recall, the context signals are re-presented to both the phoneme and word nodes. For real words, activation at the phoneme level is given a large boost by the appropriate word node, allowing phonological errors to occur at a much lower level in word than nonword recall. Because this model proposes parallel serial order mechanisms at the level of phonemes and items, lexical knowledge is represented independently from phonological STM and could be impaired separately.

Hartley and Houghton's (1996) model also adopts a similar architecture to that proposed by Burgess and Hitch, but incorporates syllable position constraints, unlike the model of Glasspool (1995). These constraints allow the model to simulate the fact that when phonemes migrate between list items (in the recall of nonwords or semantically degraded words), the syllabic position of the phonemes is typically preserved (i.e., onsets/rimes are commonly exchanged between list items, but onsets are rarely substituted for rimes). Hartley and Houghton's model represents verbal inputs at the level of syllables and phonemes. There are two pathways for encoding the phonology of syllables: the content pathway, in which syllables are linked directly with their constituent phonemes, and the structural pathway, which operates via a syllable template. This template, which represents knowledge of syllable structure, is used to assign an appropriate syllable position to each incoming phoneme. During recall, syllable nodes reactivate phonemes via the two pathways. Phoneme activation rarely reaches the threshold for recall unless both the content and structural pathways contribute. Therefore, the strongest competitors of target phonemes are the phonemes from the same syllabic position in adjacent items. Although this model does not incorporate a mechanism to account for the lexicality effect, Hartley and Houghton suggest this could be accomplished by adding familiar syllable nodes to the model.

More recently, Brown, Preece and Hulme (2000) have proposed a model of serial order memory (OSCAR) that utilises dynamic oscillators to provide a constantly changing temporal context with which items can be associated. This model is not specific to verbal STM, as the temporal context can be associated with both verbal and nonverbal events and can operate across a wide variety of delays. The authors argue, however, that dynamic oscillators could provide the temporal context signal required in models like that of Burgess and Hitch.

Page and Norris (1998) proposed a model of phonological STM that employed a rather different mechanism to underpin memory for serial order. Instead of using Hebbian learning to associate items with content units, this 'primacy model' incorporates an activation gradient across the nodes representing list items. The most active nodes, corresponding to the earliest presented items, are selected for recall first. This mechanism for representing serial order, which operates independently of the language system, is not sufficient to account for phonological similarity effects or the involvement of lexical and semantic codes in verbal STM. In order to account for these findings, the model incorporates a second stage of processing that the authors likened to speech production models, i.e., Dell and O'Seaghda's (1992) interactive activation (IA) model. This network model is, therefore, an amalgam of the theoretical approaches described previously. Parts of verbal STM are seen as underpinned by the language system, whereas other parts are seen as independent of it. The model suggests that separate mechanisms underlie memory for items and their order, with stable linguistic representations playing a much greater role in retaining the identity of items than their order. This suggestion has received some empirical support (e.g., Gathercole et al., 2001; Hulme et al., 1997; Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier, 1999; Walker & Hulme, 1999) but remains controversial. However, this model, like that of Burgess and Hitch, is not able to account for phoneme order errors, as serial order is represented solely at the level of whole items.

1.3.5 A parallel debate: conceptual knowledge and working memory

This discussion has focused on the role of lexical and semantic representations at the level of single words in verbal STM tasks, in line with the empirical investigations presented in this thesis. There is an analogous debate, however, concerning the impact of conceptual knowledge on STM capacity across the verbal and non-verbal domains. As noted above, recall is considerably better for prose than for unrelated words. Similarly, experts have superior STM for meaningful stimuli in their domain of expertise. For example, Hambrick and Engle (2002) found that memory for baseball commentaries was strongly affected by participants' prior knowledge of baseball. In addition, Chase and Simon (1973) found that chess experts were better able to remember the configuration of chess pieces than novices, but this advantage disappeared when the pieces were arranged randomly on the board. This recall advantage for meaningful material is typically explained in terms of chunking (Miller, 1956).

Ericsson and Kintsch (1995) proposed the concept of long-term working memory (LT-WM) to account for the impact of meaningfulness on WM capacity. According to this theory, activated portions of LTM, which represent the end products of processing (chunks), are kept directly accessible by means of actively maintained retrieval cues, allowing LTM to act as an extension of WM in domains of expertise. Similarly, Engle and colleagues have suggested that WM capacity can be accounted for by LTM activation and the capacity for controlled attention (Engle, Kane, & Tuholski, 1999; Kane & Engle, 2000; Rosen & Engle, 1997) and Cowan (1995; 1999) has argued that three memory components contribute to WM capacity: 1) activated portions of LTM in the focus of attention, 2) activated LTM not in the focus of attention and 3) inactive portions of LTM made accessible by retrieval cues. All of these theories suggest that pre-existing long-term representations underpin the memory advantage for meaningful material and are therefore related to the view that verbal STM corresponds to the activation of stable linguistic representations (e.g., Patterson et al., 1994; N. Martin & Saffran, 1997).

In contrast, Baddeley (2000) recently proposed a new component of the WM model – the episodic buffer – to account, in part, for the role of LTM in the immediate recall of meaningful material (also see Baddeley & Wilson, 2002). The episodic buffer is an attentionally limited temporary store capable of integrating information from multiple sources by drawing heavily on executive resources, in a process akin to chunking.

According to this viewpoint, immediate prose recall relies on the effortful and controlled integration of a diversity of representations (e.g., phonological, lexical, syntactic, semantic and conceptual), from both the WM slave systems and LTM. This approach rejects the view that WM is simply the activated portions of LTM. Instead, relevant long-term knowledge is held temporarily in an integrated state by the episodic buffer store. This proposal is therefore related to the suggestion that verbal STM utilises representations that are independent of stable linguistic knowledge.

1.4 Distinguishing between the different accounts of the relationship between verbal short-term memory and language

The three theoretical perspectives reviewed above in sections 1.3.1 to 1.3.3, and the connectionist models that have grown out of them, principally vary in terms of the degree to which verbal STM is seen as being independent of stable linguistic representations. The approach of Patterson and N. Martin sees verbal STM as an emergent property of the language system and not at all independent of it. R. Martin and Barnard alternatively suggest that verbal STM and long-term linguistic representations are separable but very closely related. In contrast, the models of Baddeley and Gathercole, Hulme and Schweickert propose a greater division between verbal STM and language processing, although anticipate that the two systems will interact. The question of whether verbal STM and language processing are underpinned by independent systems is therefore crucial for distinguishing between these models. In addition to this key issue, the various theoretical approaches make different predictions about a) the effect of lexical/semantic factors on item and order errors, b) whether stable linguistic representations contribute to ISR performance throughout the task or only during recall, c) the effect of lexical and semantic factors on the shape of the serial position curve and d) the effect of word neighbourhood size on recall. Each of these topics will be discussed in turn.

1.4.1 To what extent is verbal short-term memory separable from stable linguistic representations that underlie language processing?

Neuropsychology provides the best evidence for independence between verbal STM and linguistic representations involved in language processing. If verbal STM and language are separable processes, it should be possible for patients to have a specific impairment of verbal STM in the context of intact linguistic processing. The reverse dissociation, namely, intact ISR but poor language processing should not occur because all the models described previously predict that linguistic representations play some kind of a role in ISR. Indeed, pronounced language impairments in aphasic patients are apparently always accompanied by poor ISR performance (Heilman, Scholes, & Watson, 1976; Ostergaard & Meudell, 1984).

Relatively few patients with specific deficits of verbal STM and intact language processing have been described (see Shallice & Vallar, 1990, for a review). Patient KF (Shallice & Warrington, 1970; Warrington & Shallice, 1969) was the first such case to be reported. Two other commonly cited examples are JB (Shallice & Butterworth, 1977) and PV (Vallar & Baddeley, 1984). All three of these cases had a very severe impairment of verbal STM, with span limited to one or two items. The patients performed well on tests of auditory word identification, suggesting that their ISR deficits did not stem from speech perception problems. In addition, they were markedly impaired on verbal STM tasks that did not require spoken output (e.g., probe digit, matching span or pointing tasks), indicating that their ISR impairments were not the result of speech production difficulties.

Patients displaying this pattern of deficits almost always show rapid phonological forgetting in the Brown-Peterson task, better ISR with visual than auditory presentation (the reverse of the normal modality effect) and reduced recency but normal primacy in free recall (see Shallice & Vallar, 1990), consistent with the suggestion that they have a specific impairment of a verbal STM store. In addition, patient PV (Vallar & Baddeley, 1984) showed an effect of phonological similarity but not word length with auditory presentation. With visually presented material, she showed no effects of phonological similarity, word length or articulatory suppression. These findings suggested to Vallar and Baddeley that PV had a defective but partially functioning phonological loop store, and consequently she did not use subvocal articulation in order to rehearse verbal material.

This pattern of impairments can apparently result from phonological coding deficits (see Trojano & Grossi, 1995; Trojano, Stanzione, & Grossi, 1992). Therefore, the crucial question is whether STM-impaired patients have also deficits of phonological processing or whether their impairments are specific to short-term retention. In Shallice and Vallar's (1990) review of 14 STM patients, only 3 cases (JB: Shallice & Butterworth, 1977; PV: Vallar & Baddeley, 1984; TB: Baddeley, Vallar and Wilson, 1987) are reported as having virtually normal spontaneous speech. JB had a normal pattern of pauses in her speech. Similarly, PV was able to speak fluently at a normal rate, had normal comprehension and performed well on auditory processing tasks. The other patients in the review had speech characterised by hesitancy, word finding difficulties or phonemic paraphasias.

If the same representations underlie phonological processing and STM tasks, subtle impairments of phonology may be sufficient to produce marked verbal STM deficits, as STM tasks may be particularly phonologically demanding and therefore more vulnerable to brain damage than standard language processing tasks (Allport, 1984). Mild language-processing deficits may be demonstrated in patients like JB and PV using especially sensitive and demanding tasks, or during the initial stages of impairment. Allport (1984) found that JB was impaired at a difficult phonological discrimination task requiring same/different judgements for CVC nonwords and at an auditory lexical decision task, challenging the view that he did not have any linguistic processing deficits (however, Shallice and Vallar, 1990, point out that the phonological discrimination task made considerable demands on STM). Patient PV initially experienced word-finding difficulties and made phonemic paraphasias in spontaneous speech. Although these problems disappeared during recovery and were not present at the time of Vallar and Baddeley's study, they suggest that PV may have had subtle language processing deficits.

R. Martin and Breedin (1992) attempted to circumvent these problems by comparing the performance of patient EA, who had a severe phonological STM impairment but mild phonological processing problems, with three other aphasic patients who showed similar deficits of phonological processing. The control patients showed normal effects of phonological similarity, modality of presentation and recency in ISR, unlike EA, and much higher levels of recall. Martin and Breedin took these findings as evidence for independent phonological processing and STM capacities, as EA's phonological processing problems did not seem to account for her STM impairments. It should be noted, however, that the control patients' STM performance was impaired relative to healthy participants, again suggesting an association between phonological processing deficits and impairments of verbal STM.

Recently, Belleville, Caza and Peretz (in press) reported that patient IR, who showed a pattern of abilities and deficits typical of pure verbal STM cases, was impaired at both STM and LTM tasks involving phonological processing but intact at similar tasks involving a lexical/semantic code. These results point to an association between STM and LTM abilities and a dissociation between different linguistic representations. IR showed a reduction in the influence of phonological variables (e.g., phonological similarity) on her STM performance and enhanced effects of lexical and semantic variables (e.g., semantic similarity and concreteness). She also showed false recognition of lures that were phonologically related to target words but accurate rejection of semantically related lures after a 30 second filled delay, suggesting her long-term retention of phonological but not semantic information was impaired. Romani and Martin (1999) reported a similar association between STM and LTM; patients with a semantic STM deficit had difficulty forming semantic but not phonological long-term memories, whereas patients with a phonological STM deficit showed the opposite pattern. Therefore, the neuropsychological evidence for separable STM and LTM stores is not particularly compelling when the type of the information to be stored is taken into account.

1.4.2 The influence of lexical and semantic factors on item and order errors

The redintegration and semantic binding accounts appear to make different predictions about the influence of lexical and semantic factors on item and order errors. If the phonological trace of an item is degraded, redintegration should increase the probability of recalling the whole item and its constituent phonemes correctly but should not increase the probability of recalling the item in its correct serial position. In line with this prediction, several studies have found that lexical and semantic factors affect identity but not order errors at the level of whole items (Gathercole et al., 2001; Hulme et al., 1997; Poirier & Saint-Aubin, 1995, 1996; Saint-Aubin & Poirier, 1999; Walker & Hulme, 1999). Redintegration is similarly expected to affect item but not order errors at the level of individual phonemes (Gathercole et al., 2001). Missing or incorrectly recalled phonemes can be reinstated through the redintegration process. However, as degradation of the phonological trace is assumed to be insensitive to the lexical status of items, an equal number of word and nonword phonemes should migrate. In pure word lists, the redintegration process might be expected to correct phoneme intrusion errors. If words are presented in mixed lists with nonwords, however, word phonemes that intrude into nonwords cannot be reconstructed and it should be apparent that word and nonword phonemes migrate equally often (see Chapter 5).

In contrast, the semantic binding hypothesis predicts that the strong connections between the phonemic elements of familiar words will help to prevent phoneme migration errors. According to this theory, the phonemes of words are more likely to emerge together in ISR because lexical/semantic knowledge facilitates the binding of their phonemes into coherent items. The lack of such binding is thought to cause the frequent phoneme order errors observed in normal nonword recall (Treiman & Danis, 1988) and the word recall of SD patients (Patterson et al., 1994). Connections between the phonemes of familiar words will also reduce the incidence of phoneme identity errors, as phonemes that co-occur in words will boost each other's activation. Given that both phoneme identity and order information contribute to item memory, lexical and semantic factors are again expected to have a considerable impact on the probability of recalling whole items correctly in any serial position. The semantic binding account does not make an explicit prediction about the effect of lexical and semantic variables on item order errors. As the order of phonemes also represents the order of items, however, it is possible that these variables do have an impact on memory for serial order at the level of whole items (see Chapters 5 and 6).

1.4.3 Do linguistic representations contribute to immediate serial recall throughout the task or during the recall process?

Some versions of the redintegration hypothesis predict that stable linguistic representations only play a role in verbal STM during the process of recall (e.g., Gathercole et al, 2001; Schweickert, 1993; Walker & Hulme, 1999). In contrast, approaches that view verbal STM as an emergent property of the language system (e.g., the semantic binding hypothesis, Patterson et al., 1994) do not posit any particular role for LTM during recall. Instead, linguistic representations are expected to contribute to ISR performance throughout the task by appropriately constraining phonological activation. Lexical and semantic constraints should increase the likelihood of the network settling on the right pattern of phonological activation during encoding and help the network to maintain this pattern over the time course of the task, as well as contributing to the network's ability to produce the phonological elements of words in the right order during recall.

Studies comparing the role of long-term linguistic representations in serial recall and matching span tasks have provided some support for the redintegration hypothesis. Matching span is a serial recognition paradigm: two successive lists of items are read aloud to a participant, who is required to make a same/different judgement. This task does not require overt recall and so is expected to bypass the redintegration mechanism. Consequently, the redintegration account predicts that lexical influences will be reduced or even abolished in matching span. Knott et al. (2000) examined immediate recall and matching span performance in a patient with SD. Recall was substantially better for words that were still relatively well known by the patient, compared with words that were more semantically degraded. In contrast, there was no difference in matching span for the known and degraded words, in line with the predictions of the redintegration hypothesis.

Normal subjects also show little or no effect of lexical and semantic factors on matching span, despite showing effects of these variables in ISR. Gathercole et al. (2001) found the effect of lexicality was markedly reduced (but still significant, especially for longer lists) in serial recognition compared with serial recall in healthy children and undergraduates. In a similar vein, Thorn and colleagues (Thorn & Gathercole, 1999; Thorn, Gathercole, & Frankish, 2002) studied the serial recall and recognition performance of bilinguals in their first and second languages. A sizeable first-language advantage was obtained in serial recall but not recognition, suggesting that language-specific knowledge enhanced ISR performance during the process of recall. In addition, Walker and Hulme (1999) found no effects of concreteness in matching span performance, despite finding significant effects of this variable in ISR.

Although the semantic binding hypothesis suggests that lexical and semantic variables contribute to both recall and matching span tasks, it is not necessarily incompatible with these findings. The matching span tasks used in these studies may have been minimally sensitive to the role of lexical and semantic variables as they required changes in item order to be detected but did not require memory for the items themselves. In contrast, several studies have suggested that, at least at the level of whole items, lexical and semantic variables predominantly affect item rather than order errors (Gathercole et al., 2001; Hulme et al., 1997; Poirier & Saint-Aubin, 1995; Walker & Hulme, 1999). Consequently, the matching span tasks employed by Gathercole et al., Walker and Hulme and Knott et al. may have been relatively unaffected by lexical and semantic factors either 1) because they bypassed a redintegration process operating specifically at recall or 2) because they did not tap the processes supporting the maintenance of item information in STM. Chapter 6 compares the impact of lexical and semantic variables in traditional matching span tasks and a novel matching span task requiring memory for item identity, in order to address this issue.

Event-related brain potential (ERP) evidence (Ruchkin et al., 1999) also suggests that stable lexical representations play a role throughout ISR tasks. In this study, different patterns of ERP activity were associated with span tasks involving words and nonwords, and these differences occurred during presentation of the items and throughout a retention interval as well as during recall.

1.4.4 How do stable linguistic representations affect the serial position curve?

The account of N. Martin and Saffran (N. Martin & Saffran, 1990; N. Martin & Saffran, 1997; Saffran & Martin, 1990), which views verbal STM as an emergent property of the language system, predicts that lexical and semantic factors should have their biggest impact on the early portions of the serial position curve. Following Dell and O'Seaghda's (1992) IA model of speech production, the semantic contribution to STM tasks should be quite slow, as activation needs to spread from the phonological units to the semantic units and back again. Consequently, in an ISR task, the influence of semantics should be largest for the earliest words in the list, as there has been more time for semantic effects to build up for these items. Later portions of the serial position curve should be more heavily influenced by phonological factors, which come into play more quickly.

Interestingly, the redintegration hypothesis of Hulme et al. (1997) makes the opposite prediction. According to this theory, the phonological trace decays over time and this degradation should become more severe for the final items in the list. It is proposed that degraded items will be reinstated through a process of redintegration that utilises stable linguistic representations. As a result, the influence of lexical factors is expected to be greater towards the end of the list. This theory is inconsistent with the suggestion that the recency effect is underpinned by a relatively well-preserved phonological trace for the final items (e.g., M. J. Watkins, 1977).

A number of studies support N. Martin and Saffran's proposal that semantic factors play a greater role in the early portions of the serial position curve. For example, Watkins and

Watkins (1977) found that lists containing high followed by low frequency words were recalled better than lists containing low followed by high frequency words. Using a similar method, Brooks and Watkins (1990) found that words drawn from a single semantic category were recalled more accurately than words drawn from different semantic categories, particularly when the same category words were presented at the beginning of the list. These findings are in line with several neuropsychological studies (N. Martin & Saffran, 1990; N. Martin & Saffran, 1997; R. C. Martin & Lesch, 1996; Saffran & Martin, 1990) showing that semantically impaired patients show a reduced primacy effect, whereas phonologically impaired patients show a reduced recency effect. Martin and Saffran (1997), for example, found an association in fifteen aphasic patients between semantic impairment and the ability to recall both the first word of a two-word list and the initial phonemes of a single word.

The literature is highly inconsistent however. A number of studies have found that lexical influences are larger in the recency portion of the serial position curve, in line with the predictions of Hulme and colleagues. Hulme et al. (1997) found that the recall difference between high and low frequency words increased towards the end of the list although Walker and Hulme (1999) did not obtain a similar result for concreteness. In addition, several neuropsychological studies have found little difference in the shape of the serial position curve for relatively well-known and semantically degraded words (Forde & Humphreys, 2002; Knott et al., 1997). The semantically impaired patients in these studies generally showed substantial primacy effects but negligible recency effects, in contrast with the results of Martin and Saffran (1997).

There are several potential reasons for these discrepancies in results. Martin and Saffran's patients had substantial phonological as well as semantic impairments, whereas the patients studied by Forde and Humphreys and Knott et al. performed well on tests of phonological processing. Consequently, Martin and Saffran's patients had severe deficits of verbal STM and it was appropriate to test their ISR on single items or pairs of items. In contrast, the patients studied by Forde and Humphreys and Knott et al. had much higher levels of ISR and were tested on four to six items. Over the course of these longer lists, there may have been time for semantic representations to be activated for all of the list items, even if semantic activation lags behind phonological activation as predicted by Dell and O'Seaghda's (1992) model. Therefore, potentially crucial differences in timing between the production of single items and lists of items may explain why the predictions of Martin and Saffran and Hulme and colleagues are so different. These authors agree that semantic effects should be greater for items with a longer delay between presentation and recall, but they disagree about whether this delay will be greater for items at the beginning or end of a list. In the case of serial recall, there may be little difference in the interval between presentation and recall for items at the beginning and end of the list, as items at the end of the list are both presented and recalled last.

1.4.5 How does word neighbourhood size affect immediate serial recall?

The redintegration hypothesis predicts that degraded items in STM are reconstructed from lexical representations in LTM. If many words in LTM are potential candidates for the items in STM, recall should be poorer. Consequently, the standard version of the redintegration hypothesis predicts that words with large phonological neighbourhoods should be harder to recall than words with small phonological neighbourhoods, where phonological neighbourhood size is a measure of the number of words that sound similar to a target word. In contrast, models that view verbal STM as arising from the interaction of different stable linguistic representations make the opposite prediction. In the PDP approach of Patterson et al. (1994), for example, patterns of phonological activation that occur across a large number of words will be more robust and self-sustaining than more unusual patterns of phonological activation. In this model, phonological neighbours increase the likelihood of producing the phonology of a word correctly. One study has found some support for this second prediction (Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002). Words with large phonological neighbourhoods were recalled more accurately than words with smaller phonological neighbourhoods. The authors of this study proposed a new version of the redintegration hypothesis, in which similar stable phonological-lexical representations are associatively linked, in order to account for these findings.

1.5 Conclusions

This chapter has reviewed evidence indicating that stable linguistic representations make a contribution to verbal STM at a multitude of levels; sublexical, lexical, semantic, syntactic, thematic (see section 1.2). There are a number of accounts of the relationship between linguistic representations and verbal STM (see section 1.3) and the key difference between them is the degree to which they consider verbal STM to be a separable cognitive resource, independent from language processing. The different theories make conflicting predictions in a number of areas, although the evidence, reviewed in Section 1.4, remains largely inconclusive. Many of these areas of contention are addressed in later parts of this thesis, including the effect of stable linguistic representations on different error types (Chapter 5; see also Chapters 2, 3, 4), on portions of the serial position curve (Chapter 2; see also Chapter 6) and on recall vs. recognition tasks (Chapter 6).

It should be noted that while particular aspects of verbal STM appear to draw heavily on linguistic representations, some of the mechanisms involved in ISR might be independent of the language system. Serial recall often involves remembering novel conjunctions of familiar items (as in digit span) and it has been proposed that this aspect of ISR is not underpinned by the activation of pre-existing representations, but rather by the formation of new episodic links between activated items (e.g., Cowan, 1995). Prefrontal mechanisms underpinning attentional control are also likely to make an important contribution to ISR. The importance of executive abilities was emphasised in the original WM model (Baddeley, 1986; Baddeley & Hitch, 1974) by the inclusion of a 'central executive' subsystem. Similarly, the models of Cowan (1995; 1999) and Engle and colleagues (Engle et al., 1999; Kane & Engle, 2000; Rosen & Engle, 1997) place particular emphasis on the role of attention in STM. In addition, there may be a domain general mechanism underpinning temporal judgements that is involved in serial order memory (Brown et al., 2000). The issue of which aspects of verbal STM are dependent on stable linguistic representations and which are not arises several times throughout the thesis, most notably in Chapters 5, 6 and 7.

1.5.1 An overview of the data chapters

Chapters 2, 3 and 4 examine the impact of the loss of semantic representations on verbal STM in patients with SD. As noted above, SD patients make more frequent phonological errors in ISR for words they no longer fully understand, compared with words that they understand relatively well, supporting the view that semantics makes a major contribution to the stability of phonological representations in ISR tasks (Patterson et al., 1994). Several studies have failed to observe this recall difference between known and degraded words, however, challenging this view (see McCarthy & Warrington, 2001). In Chapter 2, the recall of known and degraded words is examined in four patients with SD. The main focus of the chapter is on methodological factors which could account for the discrepancy in the results of previous studies. It is argued that the use of small set sizes in particular could account for many of the failures to observe superior recall for known words. Chapter 3 investigates the recall of number and non-number words in the same four patients. They displayed both better ISR and comprehension of the number words, suggesting that the superior recall of numbers may be a special case of the known-degraded difference.

Although Chapters 2 and 3 support the view that semantic representations contribute to ISR, one patient failed to exhibit a known-degraded recall difference and also showed some evidence of superior phonological abilities relative to the other patients. Therefore, it is not possible to reject, on the basis of this evidence alone, the view that semantics only contributes to ISR when phonological abilities are also compromised (McCarthy & Warrington, 2001; see discussion in section 2.9). The work presented in Chapter 4 set out to investigate the phonological abilities of a larger group of SD patients more closely. ISR differences were observed in every patient, regardless of whether subtle phonological processing deficits were also detected.

Chapters 5 and 6 explore methodologies in which the performance of healthy participants mirrors that of SD patients, allowing an investigation of the role of lexical and semantic factors in normal verbal STM. Chapter 5 looks at the effect of lexical and semantic variables on the occurrence of phoneme migration errors in normal ISR. Chapter 6 examines the influence of these factors in matching span, in both SD patients and normal participants.

2 When does word meaning affect immediate serial recall in semantic dementia?

2.1 Introduction

As noted in Chapter 1, there is considerable debate about the extent to which phonological STM is functionally independent from stable lexical and semantic representations. The autonomy of the phonological STM store was questioned by Patterson et al. (1994) and Martin and Saffran (1997), who argued that lexical and semantic representations provide an important source of constraint on short-term phonological activation, particularly in demanding tasks like immediate serial recall (ISR). According to these interactive theories, there are strong connections between the phonemes representing a particular word and the lexical/semantic representations of that word, and as a result, top-down lexical/semantic activation increases the likelihood that the phonological elements of words will be produced in the correct configuration in ISR. Other authors have maintained that the integrity of phonological representations is not dependent on input from the semantic system (McCarthy & Warrington, 2001).

Some of the best evidence for a major semantic contribution to phonological coherence in verbal STM is provided by studies of patients with semantic dementia (SD), who show a specific and progressive decline in semantic memory. SD patients almost never produce phonological errors in spontaneous speech, have intact digit span and generally perform well on phonological tasks like minimal pair discrimination (Knott, Patterson, & Hodges, 1997). In contrast, it appears that all SD patients show a pattern of phonological breakdown in ISR, in which phonemes migrate to new positions in the list (McCarthy & Warrington, 1987; Patterson et al., 1994). Several studies have reported that these phonological errors occur more frequently for words that are no longer fully comprehended, compared with words that are understood relatively well (Knott et al., 1997; Knott, Patterson, & Hodges, 2000; Patterson et al., 1994). ISR differences between relatively known and degraded words have also been reported in semantically impaired patients following cardiovascular accident (Forde & Humphreys, 2002) and herpes simplex encephalitis (Caza, Belleville, & Gilbert, 2002). In addition, normal participants show a similar pattern of phonological migration errors in their serial recall of non-words that by definition lack lexical and semantic representations (Treiman & Danis, 1988).

Although ISR differences between known and degraded words do occur, there have been some notable failures to find such differences (Funnell, 1996; Lambon Ralph & Howard, 2000; McCarthy & Warrington, 1987, 2001; Warrington, 1975), and their interpretation remains controversial (see Table 2.1). Knott et al. (1997) did not find a significant recall difference between known and degraded words in one patient (BM), despite finding a difference in a second patient (AM). In addition, McCarthy and Warrington's (2001) patient MNA was able to recall a normal number of words that she did not understand. These findings appear to challenge Patterson et al.'s (1994) assertion that semantics plays a major role in maintaining the phonological coherence of words in STM. As an alternative, McCarthy and Warrington argued that verbal STM could operate without the involvement of semantics, and that additional phonological-lexical impairments were responsible for the known-degraded recall differences observed in some studies. It is important to note, however, that the patients who failed to show a difference in recall accuracy between known and degraded words still made an abnormal number of phonological errors in ISR. Moreover, the inconsistency in the size of the recall accuracy difference between known and degraded words could be a consequence of discrepancies in methodology. Consequently, the research presented in this chapter investigated the effect of various methodological factors on the size of the known-degraded recall difference in SD patients.

This work focused on four methodological variables that could influence the size of the known-degraded recall difference: 1) the method used to classify items as known and degraded, 2) the length of the lists to be recalled, 3) the frequency matching of known and degraded words and 4) the total number of known and degraded words in the lists (set size). Each of these factors will be discussed in turn.

Table 2.1: *Previous studies that have examined immediate serial recall of known and degraded words in SD patients*

Study	Patient	Set size	Frequency matched	Known - degraded difference
Patterson et al. (1994)	JL	60	No	$p < 0.001$
Patterson et al. (1994)	PP	36	No	$p < 0.001$
Patterson et al. (1994)	FM	36	No	$p < 0.001$
McCarthy and Warrington (2001)	MNA	30	Yes	n.s.
Knott et al. (1997)	AM	24	Yes	$p < 0.01$
Knott et al. (1997)	BM	24	Yes	$p = 0.09$
Knott et al. (2000)	FM	20	Yes	$p < 0.001$
Warrington (1975)	AB	15	No	n.s.
Warrington (1975)	EM	15	No	n.s.
McCarthy and Warrington (1987)	NHB	12	No	n.s.
Howard and Lambon Ralph (2000)	IW	10	Yes	n.s.
Funnell (1996)	EP	7	Yes	n.s.

The studies are arranged according to set size.

2.1.1 Method used to obtain known and degraded words

Previous studies have used a wide variety of methods to select known and degraded items for recall, including picture naming, word-picture matching, definitions, verbal fluency and spontaneous speech. As the semantic degradation underlying the known-degraded distinction varies continuously, the point of cut-off between 'known' and 'degraded' items may differ across these methods. When particularly difficult semantic tests are used, items that are selected as 'degraded' may still be understood to a certain extent, whereas when easier tests are used, items that are selected as 'known' may have lost the finer nuances of their meaning. Consequently, the method that is adopted to select the known and degraded words may affect the size of the comprehension difference between them. The choice of method may be especially critical for patients who are particularly impaired at certain types of tests. For example, Knott et al.'s (1997) patient BM was particularly poor at pictorial tasks, and the tasks used to select his known and degraded words were pictorial in nature. Consequently, his 'degraded' words may have been relatively well known, at least as assessed by verbal tests. In the present study, two different methods for obtaining known and degraded items were compared, providing some measure of the sensitivity of the known-degraded recall difference to this variable.

2.1.2 List length

SD patients almost never produce phonological errors in spontaneous speech and only occasionally produce them in single word repetition, perhaps because the STM system is not sufficiently taxed by these tasks. Consequently, longer lists may increase the likelihood of phonological breakdown for degraded words. On very long lists, however, phonological errors may fall away if few of the correct phonemes are maintained until output. Knott et al. (1997) partly attributed the small number of phonological errors made by BM to list length. He was tested on six-word lists, and largely made omission errors. Previous studies have not manipulated list length systematically, and have generally tested different patients at a single list length, set according to their word span. In this study, list length was manipulated for each patient, allowing an investigation of this factor on the occurrence of phonological errors for known and degraded words.

2.1.3 Frequency matching

Patterson et al. (1994) obtained large known-degraded recall differences for three patients but they did not match the items for word frequency. No control data were reported, making it difficult to gauge how much of the recall difference was due to the known and degraded status of the words and how much corresponded to the standard frequency effect observed in normal performance (e.g. Hulme et al., 1997). Although it is clearly problematic not to match for frequency, this process has its own inherent difficulties. First, frequency matching may use up some of the natural variation in the known-degraded dimension, as the two factors are correlated and lower frequency items generally degrade earlier in the course of the disease (Funnell, 1995). By matching closely for frequency, therefore, one is unable to maximise the known vs. degraded difference. Secondly, as known words typically have higher frequencies than degraded words, frequency matching will only be possible for a small proportion of items, leading to small set sizes. Thirdly, frequency matching can largely be achieved for medium frequency words only. Finally, there are likely to be personal oddities in word frequency, as words that individuals use regularly because of their occupations or interests will have higher personal frequencies than the database counts. Therefore, 'frequency matched' pairs may not be genuinely matched and instead may be governed by these personal oddities. In this study, the outcomes of experiments that did and did not match for frequency were compared.

2.1.4 Set size

Funnell (1996) failed to find an ISR difference between known and degraded words when there were seven words in each category and suggested that the small size of the word pool might have accounted for this null result. Some support for this suggestion is provided by Table 2.1, which lists previous studies that have examined the recall of known and degraded words in SD patients, arranged according to set size. It is clear that studies involving larger sets of known and degraded words obtained significant known-degraded recall differences more often than those involving smaller set sizes (four out of six studies vs. one out of six). There is in fact a significant correlation of 0.73 between set size and study outcome (p = 0.007). In experiments with small set sizes, the same items are presented repeatedly, making them easier to identify, retain and produce at recall. In line with this suggestion, ISR is higher in normal participants when the items on each trial are drawn from a small pool and presented repeatedly (Coltheart, 1993; Conrad, 1963). Set size may affect the recall of degraded words to a greater extent than known words because patients can become more familiar with the phonological forms of degraded words as they are repeated, allowing them to catch up with the recall of known words. Roodenrys and Quinlan (2000) found larger frequency effects for healthy participants when they were tested in an open set condition in which items were never repeated, suggesting that lexical and semantic factors might play a diminished role in verbal STM when set size is small. Contrary to this suggestion, however, Knott et al. (1997) found that while set size affected recall accuracy in their patient AB, the effect of imageability on ISR did not differ for small and large set sizes. The work presented here examines the effect of set size on the recall of known and degraded words.

2.2 Case descriptions

This work examined four SD patients, EK, GT, PD and MK, who are described below in order of severity. The same patients were examined in chapters 3 and 6 (although it was not possible to include PD in the experiments presented in Chapter 6). As these investigations were conducted in parallel with those reported in this chapter (from May 2001 to January 2002), the same case descriptions apply. Chapters 4 and 5 include data from EK, GT and some additional patients. These investigations were carried out after those reported here (from June to December 2002) and consequently EK and GT were retested on the background neuropsychological tests. The case descriptions and background neuropsychological data appropriate to Chapters 4 and 5 are included within Chapter 4.

Table 2.2 provides a summary of the background neuropsychological assessment for the patients investigated in this chapter. EK was a 60-year-old right-handed woman who left school at the age of 15 and had been experiencing worsening word-finding difficulties for around five years. She was living alone and doing occasional cooking and cleaning jobs at the time of the study. An MRI scan from 2002 showed bilateral temporal lobe atrophy that was more marked in the left hemisphere. Her neuropsychological profile was dominated by a moderate impairment of semantic memory. She performed poorly on tests requiring comprehension of words and pictures; for example, word-picture matching and the Pyramids and Palm Trees test (Howard & Patterson, 1992). She was severely anomic in spontaneous speech, word fluency tasks and confrontational picture naming. Her naming errors were predominantly omissions and semantic paraphasias. In common with other SD patients, she produced surface dyslexic errors in reading aloud and surface dysgraphic errors in spelling tasks. In contrast to her marked semantic difficulties, she was well oriented in time and place, had excellent episodic memory for recent events, and had no difficulty in remembering appointments. She performed normally on tests of visual-spatial processing from the Visual Object and Space Perception battery (VOSP, Warrington & James, 1991), and she was able to produce a good immediate copy of the Rey complex figure (Lezak, 1976). Her non-verbal reasoning on the Raven's Coloured Progressive Matrices test (Raven, 1962) was normal. Her speech was fluent and syntactically well formed despite her anomia. She had intact single word phonology and she did not make phonological errors in her spontaneous speech or picture naming. She had normal spatial STM as assessed by the Corsi block tapping task, and normal verbal STM as measured by forwards and backwards digit span (Wechsler, 1987). Her word span performance, however, was characterised by frequent phonological errors similar to those described by Patterson et al. (1994).

GT, a 71-year-old right-handed male, left school aged 14 and worked as a builder and a technician in a higher education college. At the time of the study, he had been experiencing a gradual decline in his word finding and comprehension for five years. An MRI scan from 2002 showed marked bilateral circumscribed temporal lobe atrophy. His cognitive profile was similar to the description of EK above although his semantic impairments were a little more severe. He was impaired on a range of pictorial and verbal tests of semantic memory. In contrast, he was well oriented in time and space, and had intact visual-spatial skills, non-verbal reasoning abilities and memory for recent events. His speech was fluent and syntactically well formed but characterised by anomia and frequent circumlocutions, and his conversation was repetitive. He did not make phonological errors in spontaneous speech or picture naming. He had good verbal STM as measured by forwards and backwards digit span, although his word span performance was characterised by frequent phonological errors. His hearing was slightly impaired in his right ear.

Table 2.2: *Background neuropsychological scores (2001)*

Test	Max	EK	GT	PD	MK	Controls M	Controls SD
MMSE[1]	30	27	26	13*	21*	> 24[a]	-
Coloured Progressive Matrices[2]	36	33	35	25*	22*	-	-
Digit span: forwards[3]	-	6	6	7	5	6.8[b]	0.9[b]
Digit span: backwards[3]	-	7	4	5	4	4.7[b]	1.2[b]
Spatial span: forwards[4]	-	6	5	-	5	5 - 6[c]	-
Naming	64	17*	11*	4*	2*	62.3[b]	1.6[b]
Word-picture matching	64	46*	32*	17*	11*	63.7[b]	0.5[b]
PPT: Pictures[5]	52	35*	37*	26*	33*	51.1[b]	1.1[b]
PPT: Words[5]	52	36*	32*	26*	26*	51.2[b]	1.4[b]
Category fluency (8 categories)	-	18*	11*	2*	1*	113.9[d]	12.3[d]
Letter Fluency (F, A, S)	-	29	24	22	2*	44.2[b]	11.2[b]
Rey figure immediate copy[6]	36	34	34	36	30	34.0[d]	2.9[d]
VOSP: incomplete letters[7]	20	20	18	3*	10*	19.2[b]	0.8[b]
VOSP: dot counting[7]	10	10	10	10	10	9.9[b]	0.3[b]
VOSP: position discrimination[7]	20	20	20	16*	17*	19.8[b]	0.6[b]
VOSP: cube analysis[7]	10	10	10	5*	6	9.7[b]	2.5[b]

* denotes abnormal performance (i.e., more than two standard deviations below the control mean). Figures show number of items correct.

[1] Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975)

[2] Raven's Coloured Progressive Matrices (Raven, 1962)

[3] Weschler Memory Scale - Revised (Wechsler, 1987)

[4] Weschler Memory Scale – III (Wechsler, 1997)

[5] Pyramids and Palm Trees Test (Howard & Patterson, 1992)

[6] Rey figure taken from Lezak (1976)

[7] Visual Object and Space Perceptual Battery (Warrington & James, 1991)

[a] Cutoff for normal performance

[b] Control data from Bozeat et al. (2002)

[c] Normal range for age matched participants

[d] Control data from Hodges and Patterson (1995)

PD, a 73-year-old right-handed woman, left school at the age of 14 and later worked as a regional organiser for a large charity. She had an eight-year history of worsening semantic memory problems and these were very severe at the time of testing. An MRI scan from 1997 showed very marked bilateral temporal lobe atrophy that was worse in the right hemisphere, with relative preservation of more medial temporal lobe structures including the hippocampus and also evidence of some more generalised cortical atrophy. PD was near floor on a range of tests that required comprehension of pictures and words. Early in the course of the disease, she experienced particular problems with recognising objects and people, and at the time of testing, she showed poorer performance on pictorial compared with verbal semantic tests, consistent with her predominantly right-sided atrophy (Evans, Heggs, Antoun, & Hodges, 1995). Although she had been well oriented for time and place when she first presented in 1996, she was more poorly oriented at the time of testing and occasionally became lost. She also showed some impairment in visual-spatial skills and non-verbal reasoning. PD exhibited some behavioural changes, including disinhibition, which would be consistent with the disease process affecting basal frontal as well as temporal regions (Snowden, Neary, & Mann, 1996). She became increasingly difficult to test and withdrew from the study before all the experiments reported here were completed.

MK, a 67-year-old right-handed woman, was the most severely semantically impaired patient included in the study. She left school at the age of 17 and had previously been employed in clerical work. Her family reported a three-year history of worsening semantic problems. An MRI scan from 2000 showed marked temporal lobe atrophy that was strongly lateralised to the left side. She performed at or near floor on tests of semantic memory. In contrast to her semantic impairments, she remained well oriented in time and place, and her memory for recent events was excellent. Her verbal STM was normal as assessed by forwards and backwards digit span. At the time of testing, she appeared to have good single word phonology and did not produce phonological errors in spontaneous speech or picture naming. She was impaired on tests of non-verbal reasoning and visuospatial processing, but she did not show signs of disinhibition or other behavioural changes.

2.3 Experiment 1: Immediate serial recall of frequency-matched known and degraded words defined by naming and definitions

If semantic representations make an important contribution to verbal STM. ISR should be better for words that are still relatively well understood compared with words whose meanings have become degraded. To test this prediction, known and degraded words were selected for each patient using two methods: picture naming and definitions in Experiment 1, and synonym judgement in Experiment 2. The patients were tested on a variety of list lengths in both experiments in order to investigate the effect of load on the phonological coherence of degraded words.

2.3.1 Method

The patients were asked to name eighty pictures from the Snodgrass set, as well as thirteen colours and eleven body parts, and to provide definitions for the same items. Naming attempts were considered to be correct when the patients produced the right specific label for a picture. Definitions were considered to be correct when they contained enough specific information to allow the item to be identified from its description (gestures, e.g., pointing at an item, were also accepted). For EK, GT and PD, items that were both named and defined correctly were classified as known, and items that were neither named nor defined correctly were classified as degraded. For MK, this method did not produce a sufficient number of known words and consequently, content words that she used correctly in her descriptions of eight complex pictures, including the 'cookie theft' picture, were also included.

The known and degraded words were matched for word frequency as closely as possible on an item-by-item basis using data from Celex (Baayen, Piepenbrock, & Rijn, 1993), and the MRC psycholinguistic database (Coltheart, 1981). It did not prove possible to also match every pair of words for syllable length, but the average numbers of syllables in the known and degraded words were as similar as possible. The constrained way in which words were judged to be known and degraded, and the item-by-item frequency matching of known to degraded words resulted in small set sizes for all four patients. Appendix 1 gives set size along with word frequency, length and imageability ratings for each patient's known and degraded words.

Lists of known and degraded words were assembled by selecting items at random without replacement until all the items had been used, and then repeating this process as required. The frequency-matched known and degraded word pairs were yoked so that they appeared in the same positions within corresponding lists. The patients were tested on lists containing three, four, five and six words, although PD was not tested on three word lists due to time constraints, and MK was tested on two, three, four and five words because her performance was poorer than that of the other patients. There were ten lists of known and degraded words at each length, and they were presented in a blocked fashion using an ABBA design to control for practice effects. EK and GT were tested twice on lists containing four, five and six words and MK was tested twice on lists containing four and five words, in order to increase the amount of data available for analysis. The repeated lists were separated from the original testing by a period of several weeks. Three healthy control participants were matched to each patient on the basis of sex, age and years of education. They were tested on the same lists as the patients, and also on lists up to seven words long, constructed in the same way. In this and subsequent experiments, items were read aloud at a rate of one word per second for immediate spoken serial recall.

2.3.2 Results

2.3.2.1 Recall accuracy

Both list and item recall can provide a measure of recall accuracy. The two methods generally produce the same pattern of results, although floor and ceiling effects can be less problematic for item recall. Therefore, in the interests of brevity, only item recall is reported for every experiment.

Table 2.3 shows the number of items recalled in the correct order at each list length. A series of *t* tests was used to determine if the patients performed significantly more poorly than their lowest scoring controls, combining across the different list lengths at which both patients and controls were tested. EK's performance on the known words was not impaired ($t(29) < 1$). In contrast, her recall of the degraded words was significantly poorer than that of her lowest scoring control ($t(29) = 4.23$, $p < 0.0001$). Similarly, GT's recall of the known words was actually better than his lowest scoring control ($t(29) = 2.65$, $p < 0.05$) but he was markedly impaired on the degraded words ($t(29) = 3.23$, $p < 0.01$). MK showed the largest ISR impairment, perhaps because her semantic deficits were particularly severe. Her recall was substantially impaired for both known words ($t(29) = 7.08$, $p < 0.0001$) and degraded words ($t(29) = 8.42$, $p < 0.0001$). In contrast to the other patients, PD's ISR performance was not impaired. In fact, her recall was significantly better than that of her poorest performing control, for both known words ($t(29) = 3.59$, $p < 0.001$) and degraded words ($t(29) = 2.03$, $p = 0.05$).

Two of the four patients displayed a significant recall advantage for the known words over the degraded words, consistent with the notion of a semantic contribution to verbal STM. GT showed a very substantial recall difference between known and degraded words when the data were combined across list lengths ($t(132) = 4.13$, $p < 0.0001$). MK also recalled a larger number of known than degraded words ($t(112) = 2.56$, $p < 0.05$). In contrast, no difference between the recall of known and degraded words was found for EK ($t(134) = 1.42$, n.s.) or PD ($t(54) < 1$).

None of the control participants showed superior recall of the known words, suggesting that the results obtained for GT and MK genuinely reflected the involvement of semantics in ISR and not differences in difficulty between the two sets of words. For the most part, the controls showed no difference between the known and degraded words ($t(76\text{-}98) < 1.56$, n.s., although two control participants showed superior recall of the degraded words ($t(75) = 2.48$, $p < 0.05$ and $t(78) = 2.15$, $p < 0.05$). This unexpected result may reflect the fact that the known and degraded words were not perfectly matched on every characteristic affecting verbal short-term memory and when it was not possible to find a match, the degraded items were generally selected to be easier.

Table 2.3: *The percentage of frequency-matched known and degraded words defined by naming and definitions recalled in the correct order (Experiment 1)*

	Length	2	3	4	5	6	7
EK	Known	-	90.0	78.8*	64.0	55.0	-
	Degraded	-	86.7	73.8*	59.0	50.0	-
EK controls (min)	Known	-	-	80.0	60.0	51.7	48.6
	Degraded	-	-	87.5	76.0	58.3	52.9
GT	Known	-	93.3	90.0	70.0	60.8	-
	Degraded	-	83.3	62.5*	60.0	42.5	-
GT controls (min)	Known	-	-	87.5	64.0	41.7	44.3
	Degraded	-	-	82.5	70.0	61.7	50.0
PD	Known	-	-	90.0	88.0	70.0	-
	Degraded	-	-	85.0	74.0	80.0	-
PD controls (min)	Known	-	-	82.5	58.0	51.7	42.9
	Degraded	-	-	87.5	68.0	56.7	61.4
MK	Known	70.0	70.0	57.5*	50.0*	-	-
	Degraded	65.0	60.0	43.0*	38.0*	-	-
MK controls (min)	Known	-	100	85.0	70.0	63.3	57.1
	Degraded	-	96.7	90.0	68.0	68.3	57.1
All controls (mean)	Known	-	100	91.9	74.3	62.2	57.1
	Degraded	-	98.9	91.5	80.7	70.0	61.1
All controls (SD)	Known	-	0	6.9	9.6	10.8	8.9
	Degraded	-	1.9	4.5	8.4	8.1	6.2

* denotes recall below minimum score obtained across all control participants on both known and degraded words. Min = minimum.

Table 2.3 shows that percentage recall of the known and degraded words decreased as list length was increased. List length affected the patients and control participants in a similar way, and did not systematically affect the size of the known-degraded recall difference.

2.3.2.2 Error analysis

The errors made by the patients and controls were classified as belonging to one of six categories. Omission errors occurred when fewer items were recalled than were presented. Order errors were identical to one of the target items, but were produced in the wrong place in the sequence. Repetition errors were target items recalled more than once. Intrusion errors were previously presented items recalled in the wrong list. Phonological errors contained at least half of the phonemes in a target word. Unrelated errors did not fit into any of the previous categories, and were largely accounted for by patient responses that did not overlap sufficiently with the target word to reach the criteria for a phonological error.

Table 2.4 indicates the proportion of errors in each of these categories for known and degraded words, combining across list length. Chi-square was used to determine whether the pattern of errors varied across the known and degraded words. This is standard practice in the literature but it should be noted that the test assumption of independence might be violated. There were far more errors in the phonological and unrelated categories for the patients compared with the controls. Significantly different types of errors occurred on the known and degraded words for GT ($\chi^2(5) = 27.62, p < 0.0001$), EK ($\chi^2(5) = 28.93, p < 0.0001$) and PD ($\chi^2(5) = 13.41, p < 0.01$). The standardised residuals were particularly high for phonological errors (range = 1.7 to 3.3), suggesting that this error category made a major contribution to the chi-square outcome. In contrast, MK did not make significantly different types of errors on the known and degraded words ($\chi^2(5) = 8.24$, n.s.), perhaps because a substantial number of phonological errors occurred in her recall of the known words as well as the degraded words.

Table 2.4: *Errors on frequency-matched known and degraded words defined by naming and definitions (Experiment 1)*

	Phonological		Unrelated		Omission		Order		Repetition		Intrusion	
	Known	Degrad	Known	Degrad	Known	Degrad	Known	Degrad	Known	Degrad	Known	Degrad
EK	.09*	.38*	.04*	.05*	.41	.28	.28	.19	.09	.06	.10	.05
GT	.27*	.59*	.10*	.10*	.42	.20	.09	.02	.07	.06	.05	.03
PD	.26*	.67*	.03*	.03	.19	.03	.16	.06	.32	.15	.03	.06
MK	.64*	.65*	.16*	.24*	.08	.06	.03	.00	.03	.03	.05	.02
All controls (max)	.04	.06	.02	.04	.81	.64	.43	.48	.29	.26	.37	.34

The errors in each category are expressed as a proportion of the total number of errors across all list lengths.

* denotes patient scores that were larger than the maximum observed for controls.

Fig. 2.1: *Phonological and non-phonological errors on known and degraded words defined by naming and definitions as a function of list length (Experiment 1)*

Fig. 2.1a: EK's phonological and non-phonological errors as a function of list length.

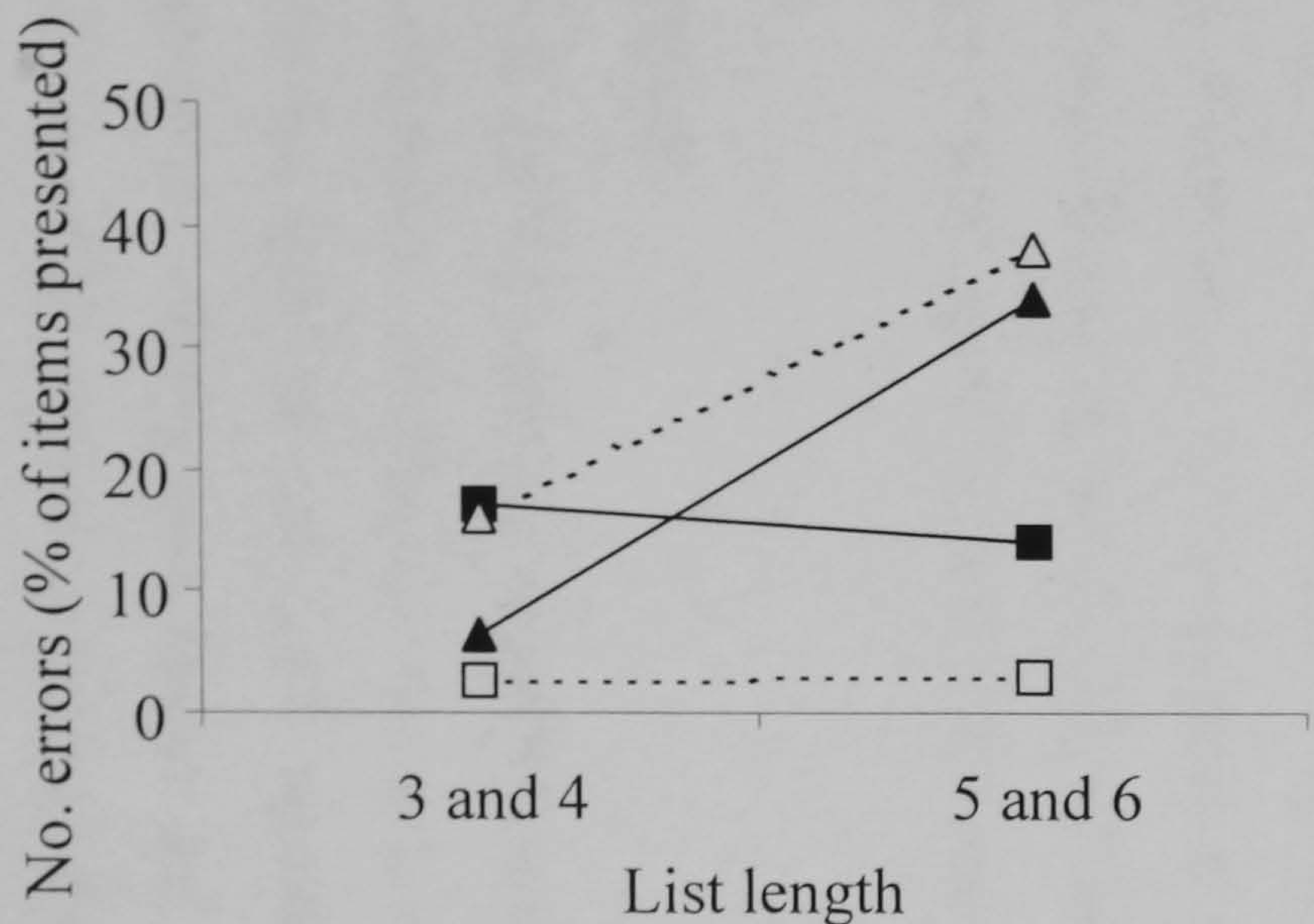


Fig. 2.1b: GT's phonological and non-phonological errors as a function of list length.

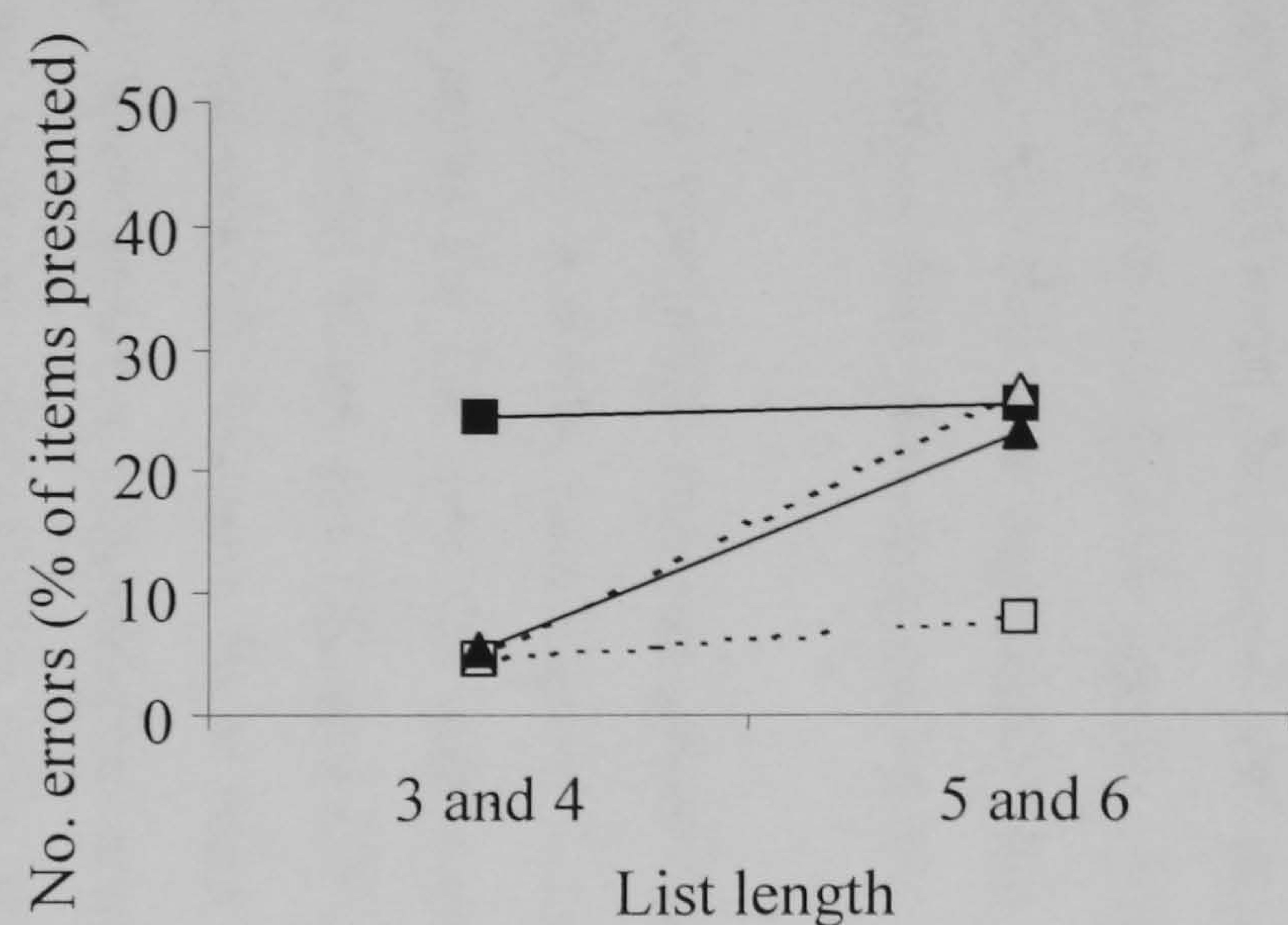
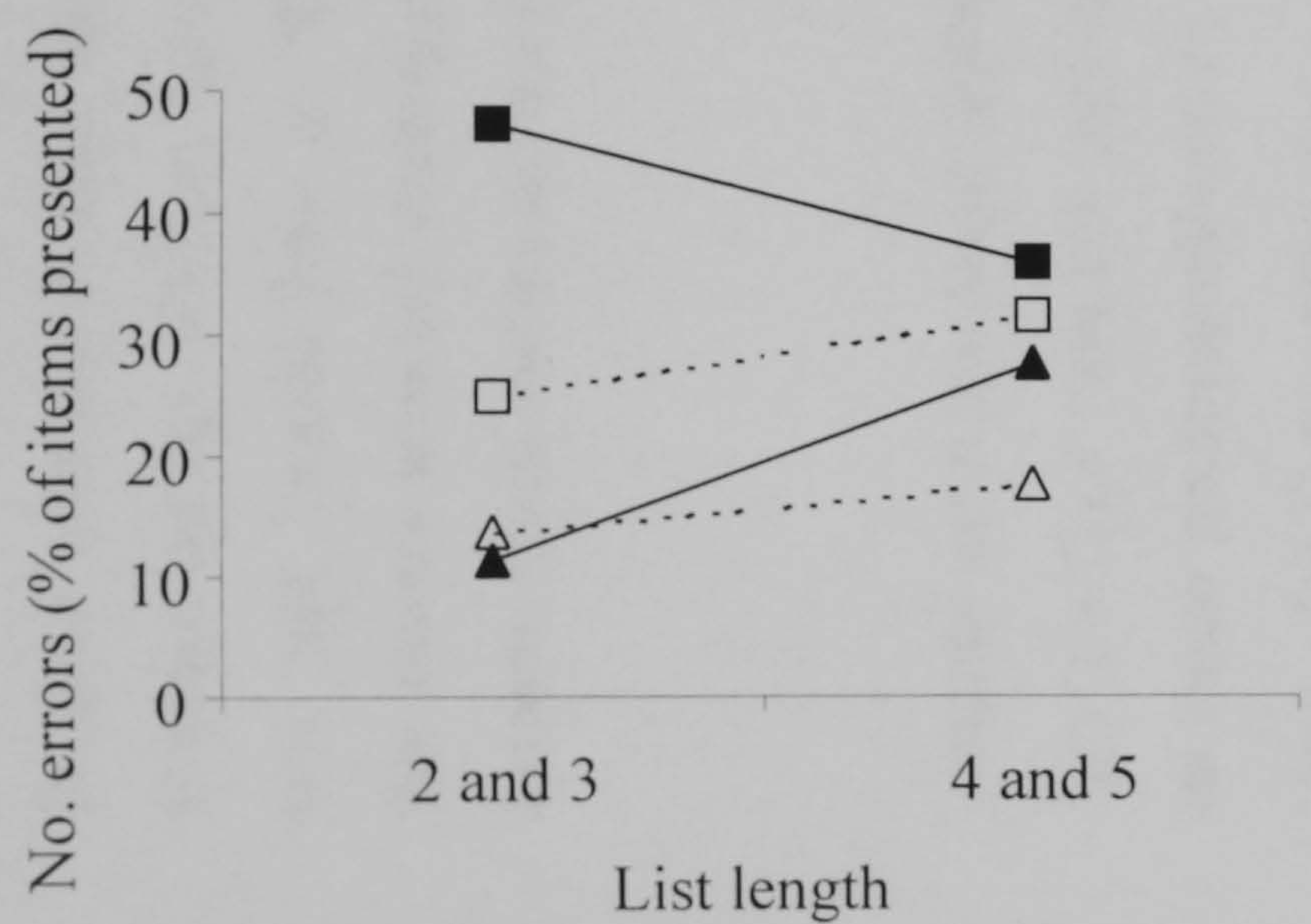
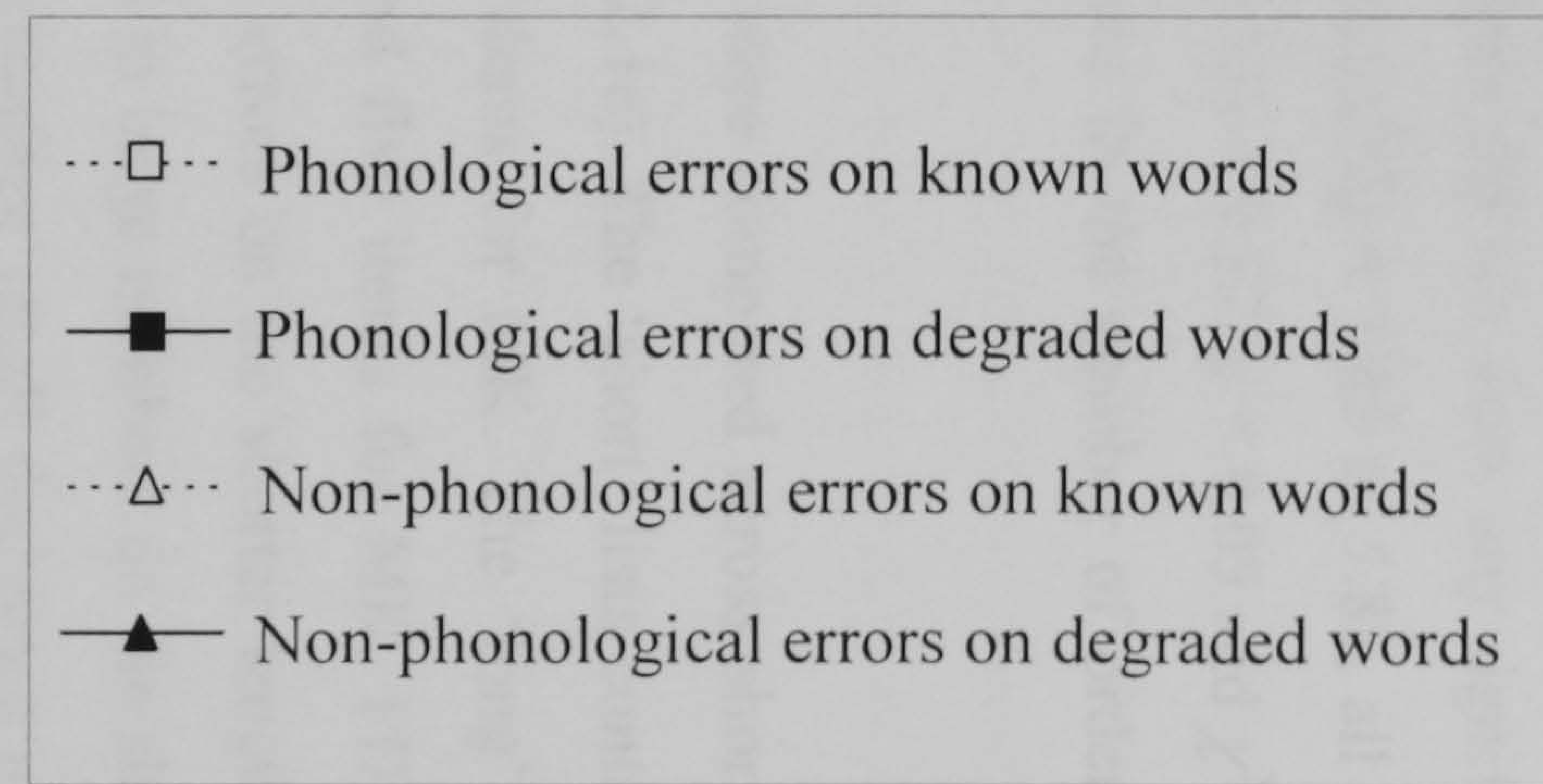


Fig. 2.1c: MK's phonological and non-phonological errors as a function of list length.





In contrast with the patients, the control participants did not make a larger number of phonological errors in their recall of the degraded words, and made very few phonological errors in either condition. Ten of the twelve control participants did not show any significant difference in errors between the known and degraded words ($\chi^2(5) = 1.85$ to 5.85, all n.s.). Two control participants did show a significant difference ($\chi^2(5) = 9.57, p < 0.05$ and $\chi^2(5) = 11.30$, $p < 0.05$), but this was apparently due to differences in the number of order and intrusion errors, rather than phonological errors.

The numbers of phonological and non-phonological errors were compared across short and long lists for EK, GT and MK (see Figures 2.1a, 2.1b and 2.1c). The 'short' lists contained three and four items for EK and GT and two and three items for MK. The 'long' lists contained five and six items for EK and GT and four and five items for MK. PD was excluded from the analysis because there were insufficient errors on the shorter lengths to analyse. For all three patients, phonological errors appeared in large numbers on the shorter lists and did not increase substantially with list length. Percentage recall decreased as list length was increased, largely because the number of non-phonological errors (predominantly omissions) rose sharply. Consistent with this pattern, phonological errors accounted for a greater proportion of the total errors on short compared with long lists, in degraded word recall, for EK ($\chi^2(1) = 15.24, p < 0.0001$), GT ($\chi^2(1) = 7.90, p < 0.01$) and MK ($\chi^2(1) = 6.75$, $p < 0.01$). The difference in the proportion of phonological to non-phonological errors on short and long lists did not reach significance for known words (EK and MK: $\chi^2(1) < 1$; GT: $\chi^2(1) = 1.93$, n.s.), presumably because the number of phonological errors was much smaller.

2.3.3 Discussion

Two patients, EK and GT, showed impaired recall of degraded but not known words, relative to control performance. GT showed the predicted recall difference between known and degraded words but this did not reach significance for EK. A third patient, MK, was markedly impaired at recalling both known and degraded words, although she nevertheless showed a significant recall difference between them. MK was the most semantically impaired patient in this study and her comprehension of the 'known' words may have been substantially impaired, although still superior to her comprehension of the degraded words. In contrast, PD's recall accuracy was at a normal level for both known and degraded words. despite her severe semantic impairments. In this respect, she was similar to McCarthy and Warrington's (2001) patient, MNA, who was characterised as 'repeating without semantics'. An analysis of the errors made by the patients and controls, however, revealed that while PD's accuracy remained at a normal level, the errors that she made were anything but normal. The number of phonological errors was much larger in the patients compared with controls, particularly for degraded words, suggesting that semantic impairment does affect the phonological coherence of items in STM.

Recall declined as list length was increased, but declined at a similar rate for known and degraded words. The size of the known-degraded difference did not vary consistently with length, as long as recall was off floor and ceiling. However, list length did appear to affect the proportion of phonological to non-phonological errors. Non-phonological errors, predominantly omissions, increased with length for both known and degraded words but phonological errors occurred frequently for degraded words, even on very short lists, and did not increase markedly with length.

2.4 Experiment 2: Immediate serial recall for frequency matched known and degraded words defined by synonym judgements

In this experiment, a second set of known and degraded words was selected using a synonym judgement task, in order to establish if the results of the first experiment would replicate regardless of the change in the method used to select the items.

2.4.1 Method

A synonym judgement task was used to produce lists of known and degraded words for EK, GT and PD. Participants were asked which of three words was closest in meaning to a target word (for example, "which word is closest in meaning to suffix: inflection, temerity or perpetrator?"). The test was administered twice on two separate occasions. Known items were selected from the consistently correct trials, and degraded items were defined as those trials where performance was consistently incorrect. MK was not included as she performed too poorly on the synonym judgement task to produce enough known words.

The frequency of the known and degraded words was matched on an item-by-item basis and the groups were matched for word length, as described for Experiment 1. Again, the set sizes were small for all four patients. Appendix 2 gives set size along with mean frequency, length and imageability for the known and degraded words selected for each patient. Lists of known and degraded words were assembled in the same way as for Experiment 1. GT and EK were tested on lists containing two, three, four and five items. PD was tested on three to five item lists. The control participants were tested on three, four, five and seven item lists. There were ten lists of known and degraded words at each length, and they were presented in a blocked fashion using an ABBA design.

2.4.2 Results

2.4.2.1 Recall accuracy

Table 2.5 shows the percentage of items recalled in the correct order at each list length by the patients and controls. In general, these words were recalled more poorly than the words in Experiment 1, perhaps because they were more abstract in nature.

A series of t tests was used to determine if the patients performed significantly more poorly than their lowest scoring controls, combining across the different list lengths at which both patients and controls were tested. GT's recall of the known words was not impaired ($t(29) = 1.55$, n.s.) but his recall of the degraded words was substantially impaired ($t(29) = 6.94$, $p < 0.0001$). EK's recall was impaired for both known words ($t(29) = 4.87$, $p < 0.0001$) and degraded words ($t(29) = 5.22$, $p < 0.0001$). In contrast, PD's recall was not impaired for either known words or degraded words ($t(29) < 1$), as in Experiment 1.

GT recalled the known words much more accurately than the degraded words ($t(78) = 4.02$, $p < 0.0001$) but there was no difference in recall between the known and degraded words for

EK ($t(77) < 1$) and PD ($t(58) < 1$). Therefore, this experiment replicated the findings of Experiment 1; the same patients showed a significant known-degraded difference, even when the words were selected using a different method. None of the controls showed a significant difference between the known and degraded words (all $t(77\text{-}98) < 1.39$, n.s.).

Table 2.5: *The percentage of frequency matched known and degraded words defined by synonym judgements recalled in the correct order (Experiment 2)*

	Length	2	3	4	5	7
EK	Known	100.0	80.0*	37.5*	48.0*	-
	Degraded	95.0	60.0*	42.5*	36.0*	-
EK controls (min)	Known	-	100.0	80.0	66.0	34.3
	Degraded	-	93.3	77.5	64.0	42.9
GT	Known	85.0	80.0*	82.5	56.0	-
	Degraded	70.0	53.3*	52.5*	28.0*	-
GT controls (min)	Known	-	93.3	85.0	62.0	42.9
	Degraded	-	86.7	77.5	62.0	47.1
PD	Known	-	86.7	70.0	52.0	-
	Degraded	-	90.0	67.5	46.0*	-
PD controls (min)	Known	-	86.7	77.5	54.0	38.6
	Degraded	-	90.0	67.5	52.0	44.3
All controls (mean)	Known	-	97.8	88.6	72.7	47.5
	Degraded	-	95.9	85.0	74.7	54.3
All controls (SD)	Known	-	4.7	8.0	10.8	9.5
	Degraded	-	4.9	9.6	14.1	9.4

* denotes recall below minimum score obtained across all control participants on both known and degraded words. Min = minimum

2.4.2.2 Error analysis

Errors were classified as for Experiment 1. Table 2.6 indicates the proportion of errors that were omissions, order errors, repetitions, intrusions, phonological errors and unrelated errors for known and degraded words, combining across list lengths. As in the previous experiment, the patients' recall was characterised by abnormally frequent phonological errors. Unrelated errors also exceeded the normal range. In contrast to Experiment 1, none of the patients showed a difference in the errors they committed on the known and degraded words (EK, $\chi^2(5) = 2.39$; GT, $\chi^2(5) = 2.12$; PD, $\chi^2(5) = 2.52$; all n.s.). This was largely because they made numerous phonological errors on both the known and the degraded words. Ten controls also showed no difference in error types between the known and degraded words ($\chi^2(5)$ = 1.80 to 7.83, all n.s.). The difference in errors approached significance for one control ($\chi^2(5) = 9.71$, $p = 0.053$) and reached significance for another ($\chi^2(5) = 15.91$, $p < 0.01$).

The numbers of phonological and non-phonological errors occurring on the shortest two lengths (two and three words) were compared with those on the longest two lengths (four and five words), in order to determine if list length affected the types of errors that were committed. PD was excluded from this analysis, as she had not been tested on the full complement of list lengths. The results were similar to those obtained in Experiment 1 (see Figures 2.2a and 2.2b). Phonological errors again appeared in large numbers on the shorter list lengths, whereas non-phonological errors (primarily omissions) increased sharply with length. EK's phonological errors accounted for a greater proportion of the total errors on short compared with long list lengths, for both known words ($\chi^2(1) = 3.78$, $p = 0.052$) and degraded words ($\chi^2(1) = 14.19$, $p < 0.001$), but the difference was more strongly significant for the degraded words because there were more errors. GT showed the same pattern for degraded words ($\chi^2(1) = 3.99$, $p < 0.05$), but the difference did not reach significance for the known words ($\chi^2(1) = 1.61$, n.s.).

Table 2.6: *Errors on frequency matched known and degraded words defined by synonym judgements (Experiment 2)*

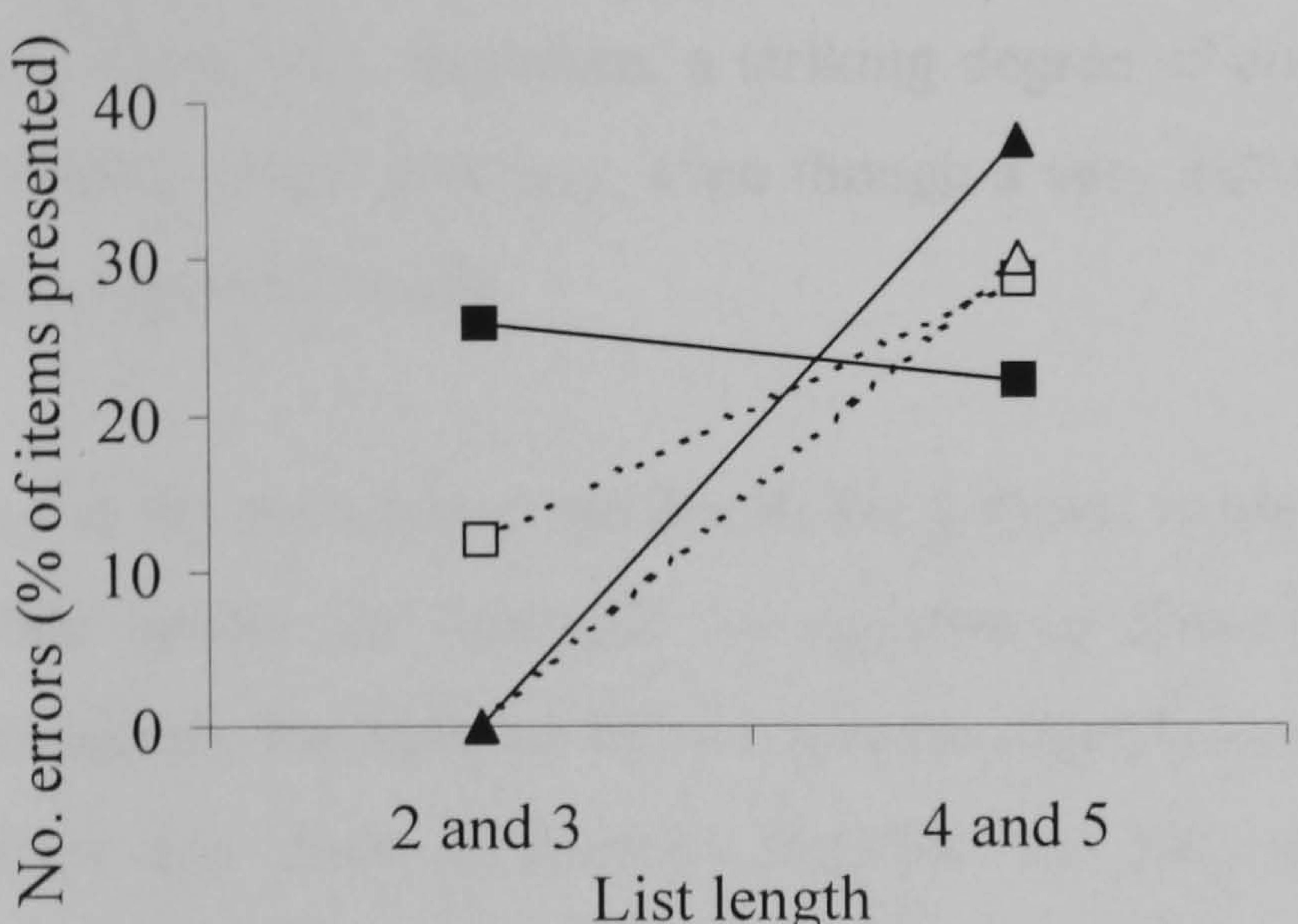
	Phonological		Unrelated		Omission		Order		Repetition		Intrusion	
	Known	Degrad	Known	Degrad	Known	Degrad	Known	Degrad	Known	Degrad	Known	Degrad
EK	.55*	.53*	.10*	.08*	.28	.35	.03	.03	.02	.00	.02	.01
GT	.66*	.59*	.11*	.15*	.24	.21	.00	.00	.00	.01	.00	.04
PD	.68*	.62*	.08*	.10*	.20	.24	.03	.02	.00	.02	.03	.00
All controls (max)	.14	.17	.06	.05	.74	.75	.32	.28	.18	.21	.37	.55

The errors in each category are expressed as a proportion of the total number of errors across all list lengths.

* denotes patient scores that were larger than the maximum observed for controls.

Fig. 2.2: *Phonological and non-phonological errors on known and degraded words defined by synonym judgement as a function of list length (Experiment 2)*

Fig. 2.2a: EK's phonological and non-phonological errors as a function of list length.



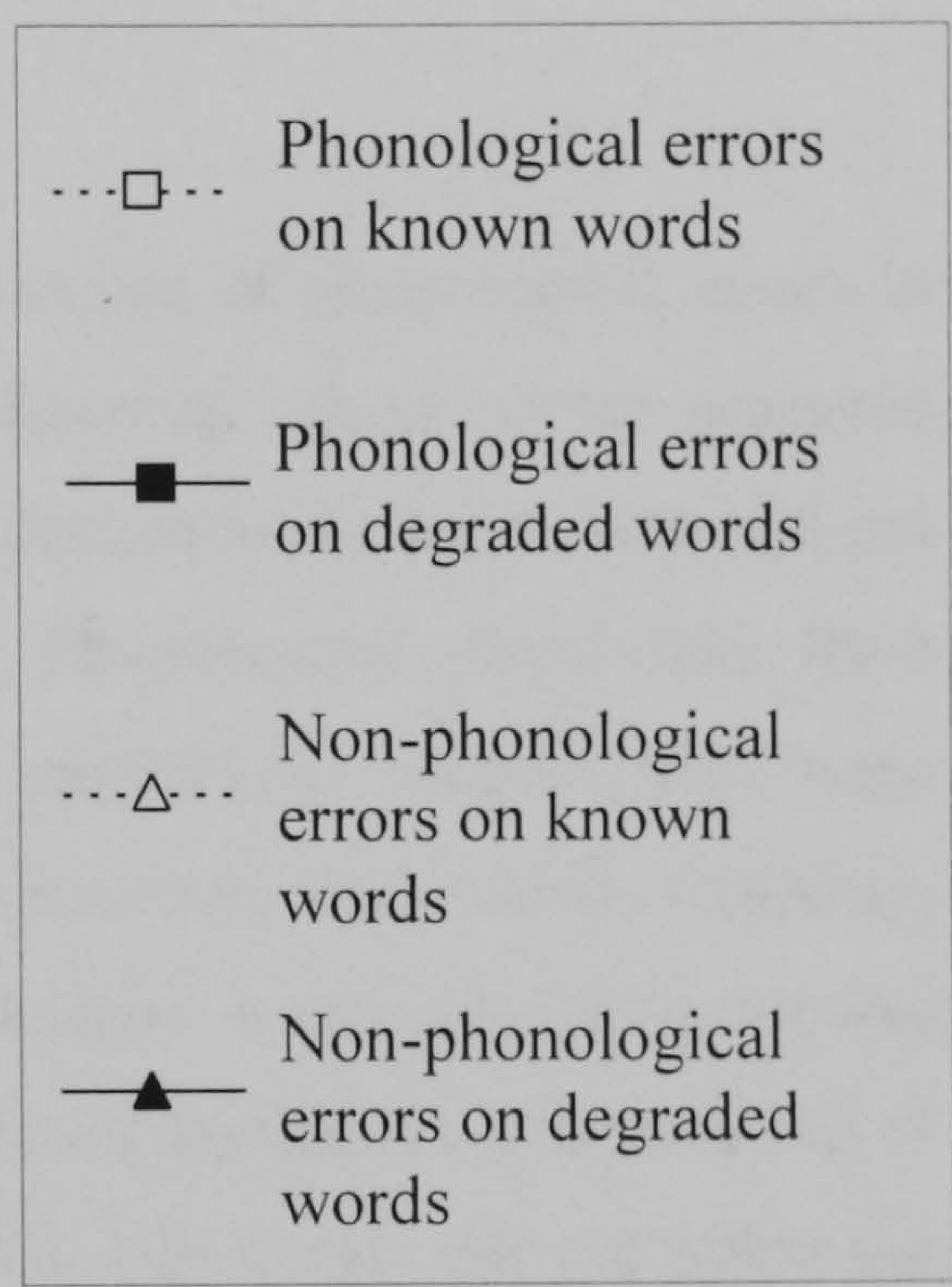
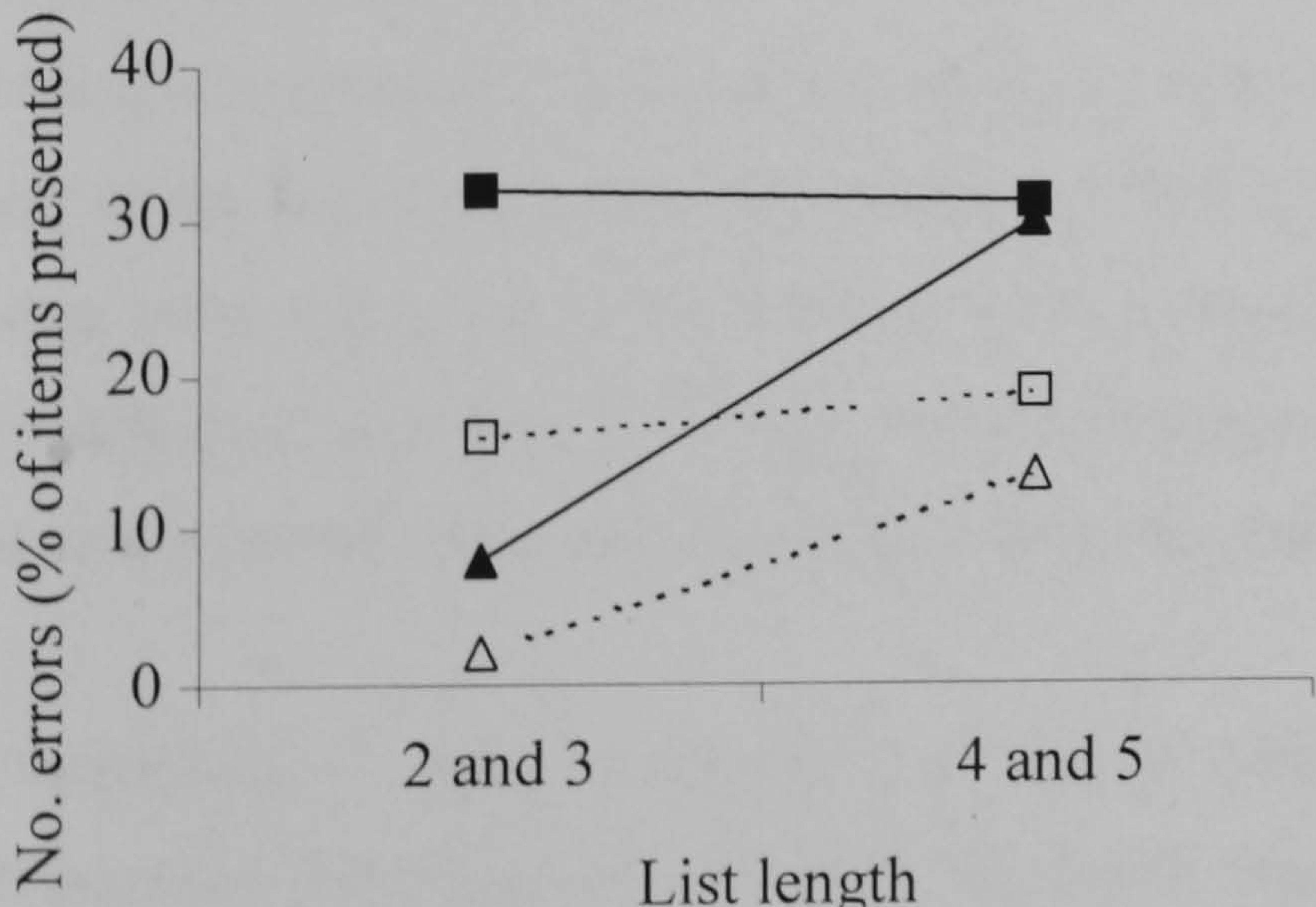


Fig. 2.2b: GT's phonological and non-phonological errors as a function of list length.



2.4.3 Discussion

As in Experiment 1, the patients' recall was impaired relative to control performance with the exception of PD, who showed normal recall accuracy despite her severe semantic impairments (though her error pattern was clearly abnormal). GT again showed a strong advantage in ISR for words that were relatively well understood, whereas EK and PD did not. There was, therefore, a striking degree of consistency between the two experiments in terms of recall accuracy, even though a very different method was used to select the known and degraded words.

As in the previous experiment, the patients made a large number of phonological errors in ISR, unlike the controls. In contrast to Experiment 1, however, these errors occurred frequently for both the known and the degraded words, and consequently the patients did not show any error differences between the two conditions. Phonological errors may have occurred more frequently for the known words in this experiment because they were generally less imageable and familiar, and may have been less well understood. Certainly, EK and GT were unable to provide definitions of all of the 'known' words selected according to synonym judgement (EK correctly defined 8 out of 15 known words, GT defined 6 out of 14 known words, PD was not tested due to time constraints). List length did not affect the size of the known-degraded difference but did affect the proportion of phonological errors to other error types, as in Experiment 1. The errors on the shorter lists were predominantly phonological in nature. On longer lists, phonological errors remained frequent but other types of errors (particularly omissions) became more common.

Experiments 1 and 2 suggest that certain patients consistently show substantial ISR differences between known and degraded words, whereas others do not. However, methodological factors could have been responsible for the failure to observe a known-degraded accuracy difference for EK and PD, especially as they both showed the predicted difference in errors in one experiment. The following experiments examined whether a significant difference between known and degraded words could be obtained for these patients in more favourable circumstances; for example, when the known and degraded words were not matched for frequency and when set size was larger. Unfortunately, PD did not wish to participate in further research, and so it was only possible to investigate the role of these factors in the recall of EK, GT and MK.

2.5 Experiment 3: Known and degraded words not matched for frequency

In the experiments above, the known and degraded words were matched on an item-by-item basis for frequency in order to ensure that superior known word recall did not result from the normal frequency effect. However, frequency matching has a number of drawbacks, outlined in section 2.1.3, which may mask genuine known-degraded recall differences. In this experiment, the patients and controls were compared on non-frequency matched known and degraded words. The known words were higher in frequency and easier to recall than the degraded words for both the patients and controls, but the known-degraded difference was expected to be much more substantial for the patients.

2.5.1 Method

Two sets of known and degraded words were selected using naming and definitions, as in Experiment 1. The words were matched on an item-by-item basis for syllable length but not for frequency, and consequently the known words were substantially higher in frequency than the degraded words. Not matching the items for frequency allowed set size to be larger in this experiment (see Appendix 3). Every participant was tested on four and five item lists. MK was additionally tested on three item lists, and the controls were tested on seven item lists. The lists were constructed following the method described for Experiment 1. There were twenty lists of known and degraded words at each length, and they were presented in blocks using an ABBA design. Six age-matched control participants were tested on the material for each patient.

2.5.2 Results

Table 2.7 shows percentage recall for the known and degraded words. Both EK and GT recalled the known words at a normal level: their recall was within the range of both the six controls tested on the same material, and all of the control data for the three patients taken together (eighteen data points from twelve participants). MK's recall of the known words was impaired, however, falling below the lowest scores obtained for controls. Recall of the degraded words fell substantially below the control range for all three patients.

All three patients recalled the known words significantly better than the degraded words when the data were combined across list lengths. EK showed a substantial advantage for the known words ($t(71) = 5.92$, $p < 0.0001$), in contrast to the results of Experiments 1 and 2. The difference was also highly significant for GT ($t(77) = 3.41$, $p < 0.001$) and MK ($t(110) = 5.71$, $p < 0.0001$). None of the controls showed a significant recall difference between known and degraded words (all $t(114\text{-}158) < 1.24$, n.s.). Therefore, the controls did not recall the known words better than the degraded words even though they were higher in frequency.

The difference between the patients and the controls is further illuminated by a comparison of the size of the known-degraded recall difference. All of the data collected for the patients was included in this analysis, but only five and seven item lists were included for the controls (three and four item lists were excluded as recall was close to ceiling, potentially reducing the size of the known-degraded difference). As it was necessary to compare patients and controls on lists of different lengths, percentage differences were contrasted. The known-degraded difference for all eighteen controls was centred on zero (mean = 0.12, range = -6.3 to 5.4). The known-degraded differences for all three patients were much larger than the maximum difference observed in the controls (EK = 38.9, GT = 18.9, MK = 24.2). The participants' errors, which followed the pattern reported for the other experiments in this chapter, are not discussed in detail here.

Table 2.7: *The percentage of non-frequency matched known and degraded words recalled in the correct order (Experiment 3)*

	Length	3	4	5	7
EK	Known	-	91.3	78.0	-
	Degraded	-	66.3*	48.0*	-
EK controls (min)	Known	-	88.8	68.0	47.9
	Degraded	-	91.3	71.0	52.1
GT	Known	-	93.8	68.0	-
	Degraded	-	78.8	46.0*	-
GT controls (min)	Known	-	80.0	63.0	40.7
	Degraded	-	78.8	55.0	43.6
MK	Known	75.0*	60.0*	68.3*	-
	Degraded	38.3*	36.7*	31.7*	-
MK controls (min)	Known	97.0	84.0	81.0	50
	Degraded	95.0	88.0	66.0	51
All controls (mean)	Known	99.2	93.3	83.4	59.4
	Degraded	98.9	93.5	81.9	60.3
All controls (SD)	Known	1.4	5.9	10.2	11.1
	Degraded	2.0	5.2	10.5	9.5

* denotes score below range for each patient's controls. Min = minimum

2.5.3 Discussion

In this experiment using known and degraded words not matched for frequency, EK showed a substantial recall advantage for known over degraded words in addition to GT and MK, in contrast to the results of Experiments 1 and 2. As the controls did not show this advantage, it seems unlikely that the patients' recall could be accounted for by the higher frequency of the known words, or by any other difference between the two sets of words. Instead, the status of the words as known and degraded appeared to be crucial. The lack of frequency matching in this experiment allowed EK's set size to be much larger than in the previous experiments, and this difference may have accounted for the discrepancy in results. The following experiments investigated the effect of set size on the recall of known and degraded words more systematically.

2.6 Experiment 4: The effect of set size on the recall of known and degraded words

The set sizes in Experiments 1 and 2 were small because of the constrained way in which the words were selected. In this experiment, the words from these experiments were pooled to increase the set size.

2.6.1 Method

The lists of four and five known and degraded words used in Experiments 1 and 2 were re-presented to EK and GT, but were interspersed so that a list from Experiment 1 (based on naming and definitions) was followed by a list from Experiment 2 (based on synonym judgements). Consequently, a larger number of items occurred between repetitions of the same word. The lists in this new large set size condition were identical to those in Experiments 1 and 2, which constituted the small set size condition, and were presented in the same way. Six age-matched control participants were tested on the same material as the patients and also on lists of seven words. The number of words in the large and small set size conditions are shown in Appendices 1 and 2 for each patient.

2.6.2 Results

2.6.2.1 Recall accuracy

Tables 2.8 and 2.9 show the percentages of words that were recalled by patients and controls in the large and small set size conditions. Table 2.8 shows the data for naming and definitions items and Table 2.9 shows the items based on synonym judgement. As the patients were tested twice on the words selected by means of naming and definitions performance in the small set size condition (Experiment 1), the data in Table 2.8 is averaged across the two presentations. EK showed a significant difference between known and degraded words in the large set size condition ($t(70) = 3.19, p < 0.01$) but not in the small set size condition ($t(78) < 1$) supporting the notion that set size can affect the magnitude of known-degraded recall differences. A more detailed analysis revealed that, in the large set size condition, the known-degraded difference was significant for synonym judgement words ($t(35) = 3.74, p < 0.001$) but not naming and definition words ($t(36) = 1.48$, n.s.). Therefore, most of the effect of the set size manipulation was brought about by the synonym judgement words, which were recalled more poorly and with a greater proportion of phonological errors than the naming and definitions words in Experiments 1 and 2. GT recalled the known words better than the degraded words regardless of whether set size was large ($t(67) = 3.92, p < 0.001$) or small ($t(76) = 4.64, p < 0.0001$). This was the case for naming and definitions words (large set: $t(29) = 2.54, p < 0.05$; small set: $t(34) = 3.07, p < 0.01$) as well as synonym judgment words (large set: $t(35) = 3.33, p < 0.01$; small set: $t(38) = 3.86, p < 0.001$).

In contrast, none of the control participants showed a significant known-degraded recall difference for either naming and definitions items or synonym judgements items in the large set size condition (all $t(53\text{-}58) < 1.67$, n.s.). This analysis combined scores on four, five and seven-word lists. The performance of controls in the small set size condition was discussed for Experiments 1 and 2 and will not be considered here.

Table 2.8: *The effect of set size on the percentage recall of known and degraded words defined by naming and definitions (Experiment 4)*

		EK	EK Controls				GT	GT Controls				All Controls			
	List length	4+5	4+5		7		4+5	4+5		7		4+5		7	
			Mean	Min	Mean	Min		Mean	Min	Mean	Min	Mean	SD	Mean	SD
Large set	Known	73.3*	85.2	78.9	50.5	50.0	84.4	89.6	78.9	66.2	57.1	87.4	8.6	58.3	10.0
	Degraded	63.3*	89.5	82.2	57.8	51.4	62.2*	88.1	82.2	62.4	52.9	90.9	5.7	57.1	8.1
Small set	Known	70.6	80.7	68.9	53.3	48.6	78.9	84.8	74.4	60.0	44.3	82.8	9.2	56.7	10.1
	Degraded	64.4*	86.7	81.1	55.7	52.9	61.1*	83.3	75.6	60.0	50.0	85.0	6.1	57.9	6.2

Table 2.9: *The effect of set size on the percentage recall of known and degraded words defined by synonym judgement (Experiment 4)*

		EK	EK Controls				GT	GT Controls				All Controls			
	List length	4+5	4+5		7		4+5	4+5		7		4+5		7	
			Mean	Min	Mean	Min		Mean	Min	Mean	Min	Mean	SD	Mean	SD
Large set	Known	56.7*	84.4	70.0	49.0	41.4	67.8*	92.6	87.8	61.4	50.0	88.5	10.0	55.2	10.3
	Degraded	31.1*	83.0	66.7	52.9	40.0	40.0*	91.1	80.0	60.0	44.3	87.0	12.0	56.4	14.6
Small set	Known	41.1*	78.9	73.3	43.8	40.0	67.8*	85.2	72.2	54.8	42.9	82.0	9.0	48.3	11.0
	Degraded	38.9*	77.8	72.2	51.9	42.9	38.9*	83.0	70.0	60.5	47.1	80.2	10.2	56.4	10.6

Note: In both tables, * denotes recall below minimum score obtained across all control participants on both known and degraded words. Min = minimum.

2.6.2.2 Error analysis

Table 2.10 shows the number of phonological and non-phonological errors (omission, order, repetition, intrusion, and unrelated errors) that occurred on the naming and definitions and synonym judgement items in the large and small set size conditions. Set size did not affect the balance of phonological to non-phonological errors produced by EK, on either the naming and definitions items (known words: $\chi^2(1) = 1.07$, n.s.; degraded words: $\chi^2(1) < 1$) or the synonym judgement items (known and degraded words: $\chi^2(1) < 1$). Set size did not affect GT's errors on the naming and definitions items (known and degraded words: $\chi^2(1) < 1$) but did affect his errors on the synonym judgement items (known words: $\chi^2(1) < 1$; degraded words: $\chi^2(1) = 4.97$, $p < 0.05$). A greater proportion of his errors were phonological in nature when set size was large.

Table 2.10: *Errors on known and degraded words as a function of set size (Experiment 4)*

			Phonological		Non-phonological	
		Set size	Known	Degraded	Known	Degraded
Naming and definitions	EK	Large	.04	.12	.23	.21
		Small	.01	.16	.29	.19
	GT	Large	.07	.21	.09	.17
		Small	.06	.26	.14	.12
Synonym judgement	EK	Large	.18	.31	.24	.37
		Small	.29	.22	.30	.38
	GT	Large	.23	.43	.09	.16
		Small	.19	.31	.13	.30

The errors in each category are expressed as a proportion of the number of items presented.

2.6.3 Discussion

Set size affected the magnitude of the known-degraded recall difference for EK but not GT. EK recalled the known and degraded words at an equivalent level in the small set size condition, but showed a significant advantage for known over degraded words in the large set size condition. EK's recall of degraded words defined by synonym judgements particularly seemed to benefit from small set sizes, perhaps because in Experiments 1 and 2 these words were more difficult to recall than those selected according to naming and definitions. For GT, a greater number of phonological errors occurred when set size was large, consistent with the notion that small set sizes improve recall by increasing the familiarity of degraded phonological forms. However, it remains unclear why set size failed to affect GT's recall accuracy. One possibility is that the variation of set size was relatively subtle in this experiment. The manipulation was applied only by changing the distance between repetitions of the same word and did not affect the total number of times each word was presented. In addition, overall set size was larger for EK than GT, so the distance between presentations of the same word may not have been large enough to produce an effect in GT. Therefore, the effect of set size was examined in a second experiment that manipulated the total number of times each word was repeated and which kept set size equal for the two patients.

2.7 Experiment 5: A second look at the effect of set size on the recall of known and degraded words

This experiment provides a replication of the previous set size findings using a rather different method.

2.7.1 Method

The methods used to select known and degraded words in previous experiments were relaxed, in order to increase the number of words that were available for testing. Known words, selected using naming and definitions and synonym judgement, were supplemented with words that were produced correctly in fluency tasks. Degraded words were identified using the procedure described for Experiments 1 and 2. The known and degraded words were matched for syllable length and frequency on an item-by-item basis using data from Celex (Baayen et al., 1993), and the MRC psycholinguistic database (Coltheart, 1981). The characteristics of these words are described in Appendix 4.

Four-item lists of known and degraded words were assembled at three set sizes: 36 words (each word was repeated twice), twelve words (each word was repeated six times) and four words (each word was repeated eighteen times; i.e., each list contained the same four items in a different order). There were three different sets of twelve words, which together made up the complete set of 36, and there were three different sets of four words, which together made up the first set of twelve words. Therefore, any recall differences between the 36-word set and the twelve-word sets were likely to be the result of the number of times each word was repeated and not the words included in each set. There were eighteen lists of known and degraded words in each set presented in blocks using an ABBA design.

Testing took place over two sessions that were several weeks apart. On the first session, the 36-word set was tested, followed by the first set of twelve words and two sets of four words. On the second session, the 36-word set was tested a second time to increase the amount of data available in that condition, followed by the remaining two sets of twelve words and the final set of four words. Three age-matched control participants were tested on the same lists as the patients, and also on lists of seven words (although they were only tested on the 36-word set once). The lists of seven words necessarily included a larger number of repetitions of each word. For this reason, the controls were not tested on lists of seven words in the four-word set condition, as most of the items would have had to be presented twice in each list.

2.7.2 Results

2.7.2.1 Recall accuracy

Table 2.11 shows the percentage of known and degraded words that were recalled at each set size. The overall pattern was similar to that observed in Experiment 4. EK showed a more substantial recall difference between known and degraded words in the 36-word set condition ($t(66) = 3.52$, $p < 0.001$) compared with the 4-word set condition ($t(92) = 2.92$, $p < 0.01$), although the advantage for known over degraded words was still significant at the smallest set size. The known-degraded difference approached significance for the 12-word set ($t(101) = 1.67$, $p < 0.1$). As in the previous experiment, smaller set sizes facilitated EK's recall of degraded words to a greater extent than known words. Set size did not significantly affect the recall of known words (36 vs. 12-word set: $t(77) = 1.61$; 12 vs. 4-word set $t(98) = 1.38$, both n.s.). In contrast, degraded words were recalled better when the set size was smaller (36 vs. 12-word set: $t(75) = 3.38, p < 0.01$; 12 vs. 4-word set: $t(105) < 1$).

As in the previous experiment, GT showed a substantial recall difference between known and degraded words at every set size (36-word set: $t(66) = 2.72, p < 0.001$; 12-word set: $t(100) = 4.19$, $p < 0.0001$; 4 word set: $t(102) = 5.99$, $p < 0.0001$). The advantage GT showed for known words actually became greater as set size was decreased, in contrast with EK, as decreases in set size enhanced the recall of known words to a greater extent than degraded words. Set size did not significantly affect the recall of degraded words (36 vs. 12-word set: $t(78) = 1.14$; 12 vs. 4-word set: $t(103) = 1.58$, both n.s.). However, known words were recalled better when the set size was smaller (36 vs. 12-word set: $t(78) < 1$; 12 vs. 4-word set: $t(104) = 3.21$, $p < 0.01$). None of the control participants showed a significant difference between the known and degraded words at any set size, collapsing across list length (36-word set: $t(69\text{-}70) < 1.14$, n.s.; 12-word set, $t(213\text{-}214) < 1$).

		EK	EK controls				GT	GT controls				All controls			
	List length	4	4		7		4	4		7		4		7	
			Mean	Min	Mean	Min		Mean	Min	Mean	Min	Mean	SD	Mean	SD
36 set	Known	70.8*	92.6	90.3	50.3	46.8	77.1	87.5	75.0	53.2	35.7	90.0	7.9	51.7	12.7
	Degraded	50.7*	92.6	90.3	51.3	44.4	61.1*	88.0	73.6	55.8	32.5	90.3	8.7	53.6	15.7
12 set	Known	78.2*	94.1	90.7	58.9	49.5	75.5*	90.3	80.6	58.4	49.5	92.2	6.9	58.6	9.4
	Degraded	70.4*	94.4	90.3	60.4	49.2	54.2*	90.0	81.5	58.0	50.3	92.2	6.5	59.2	9.7
4 set	Known	83.3*	97.8	95.4	-	-	88.9*	93.5	89.8	-	-	95.7	4.3	-	-
	Degraded	71.8*	97.8	94.4	-	-	62.5*	92.7	88.4	-	-	95.3	5.2	-	-

* denotes recall below minimum score obtained across all control participants on both known and degraded words

2.7.2.2 Error analysis

Table 2.12 shows the number of phonological and non-phonological errors (omission, order, repetition, intrusion, and unrelated errors) that were made at each set size. As in Experiments 1 and 2, a greater proportion of phonological errors occurred for the degraded words than for the known words. EK showed a significant known-degraded error difference for the 36-word set ($\chi^2(1) = 10.34$, $p < 0.01$), the 12-word set ($\chi^2(1) = 6.92$, $p < 0.01$) and the 4-word set ($\chi^2(1) = 11.58$, $p < 0.001$). GT made a larger number of phonological errors on the known words and consequently the known-degraded error difference did not reach significance for the 36-word set ($\chi^2(1) < 1$). However, it did reach significance for the 12-word set ($\chi^2(1) = 4.73$, $p < 0.05$) and the 4-word set ($\chi^2(1) = 6.48$, $p < 0.05$).

For EK, the pattern of errors changed smoothly over the three set sizes. Although there was no significant effect of set size on the pattern of errors when the known and degraded words were considered separately (for all possible comparisons between set sizes, $\chi^2(1) < 3.21$, n.s.), the difference did reach significance when the errors made on known and degraded words were combined. A greater proportion of the errors were phonological for large compared with small set sizes (36-word set vs. 4-word set: $\chi^2(1) = 6.29$, $p < 0.05$; 36-word set vs. 12-word set: $\chi^2(1) = 3.60$, $p = 0.06$; 12-word set vs. 4-word set: $\chi^2(1) < 1$). GT did not show this reduction in phonological errors with smaller set sizes. He showed no effects of set size on the balance of phonological to non-phonological errors for known words (for all comparisons, $\chi^2(1) < 1.01$, n.s.). For degraded words, there were no significant differences in errors between the 36- and 4-word sets ($\chi^2(1) = 3.29$, n.s.) or between the 12- and 4-word sets ($\chi^2(1) < 1$). There were, however, more phonological errors in the 12-word set compared with the 36-word set ($\chi^2(1) = 5.61$, $p < 0.05$).

Table 2.12: *Errors on known and degraded words as a function of set size (Experiment 5)*

		Phonological		Non-phonological	
	Set size	Known	Degraded	Known	Degraded
EK	36	.10*	.33*	.18	.15
	12	.06*	.16*	.15	.13
	4	.02*	.14*	.14*	.13*
GT	36	.17*	.26*	.07	.10
	12	.21*	.39*	.07	.05
	4	.07*	.31*	.05	.05
Controls (max): length = 4	36	.03	.04	.22	.22
	12	.02	.02	.19	.19
	4	.00	.00	.10	.11
Controls (max): length = 7	36	.06	.04	.64	.67
	12	.02	.02	.51	.51
	4	-	-	-	-

The errors in each category are expressed as a proportion of the number of items presented

* denotes patient scores that were larger than the maximum observed for controls

2.7.3 Discussion

Some interesting differences between EK and GT emerged in this experiment. EK's recall of the degraded words improved as set size was decreased and she made fewer phonological errors. She may have become more familiar with the phonological forms of the degraded words as they were repeated in the smaller set size conditions, and this would have increased the likelihood that their phonemes were produced in the correct configuration. Consequently, the magnitude of the recall difference between known and degraded words was smaller for more limited sets of words. In contrast, GT showed enhanced recall of known words when set size was small, but set size did not affect his recall of degraded words. As a result, the magnitude of the recall difference between known and degraded words was actually larger for limited word sets.

These results suggest that the differences in the effect of set size observed for EK and GT in Experiment 4 did not occur because the manipulation of set size was relatively weak. Instead, these findings point to an effect of semantic knowledge on the degree to which repeatedly presenting the phonological forms of words enhances their immediate recall. The known and degraded words may have been more semantically degraded for GT than for EK, as his semantic impairments were more severe. GT may have had very little understanding of his degraded words but some partial understanding of his known words and consequently his rate of phonological errors was consistently high throughout. In contrast, EK may have had more residual knowledge about her degraded words and substantially intact understanding of her known words. In both patients, set size may have had the most substantial impact on those words that were partially comprehended, namely, EK's degraded words and GT's known words. Some support for this hypothesis is provided by a closer look at the patients' definitions. Words were classified as 'known' in both Experiments 4 and 5 if they could be both named and defined correctly. As definitions were considered to be correct when they provided enough specific information about an item to allow it to be identified from its description, responses in the 'incorrect' category could indicate incomplete or partially correct information. For example, EK defined the word 'crocodile' as "an animal of some kind" and 'lung' as "something in your throat". These responses indicate that she had some knowledge of the general meanings of these words, even though they were classified as 'degraded' in this study.

The quality of the patients' definitions was examined quantitatively to evaluate the suggestion that EK's better recall of degraded items with smaller set sizes corresponded with partial knowledge of these items. Definitions were awarded one point when they were specific and entirely correct. Partially correct definitions were awarded half a point and no credit was given for entirely incorrect definitions or failures to respond. Definitions were available for all of the words used in Experiment 4 – both those selected according to naming/definitions and synonym judgement tests. GT defined each word only once, whereas EK was asked to define the words selected according to naming and definitions both before the start of the ISR experiments and again, several months later, when they were completed.

When two definitions were available for EK, an average of their scores was used. Definitions were also available for many but not all of the items used in Experiment 5. For the degraded items from Experiment 4, EK produced 2 correct definitions (for items selected according to synonym judgement) and 13 partially correct definitions. She failed to provide any correct information for 11 items. GT produced 2 correct definitions, 3 partially correct definitions and 18 entirely incorrect/no response errors. This difference between the patients in the number of partially correct definitions was significant ($\chi^2(2) = 7.91, p < 0.05$). Similarly, for the degraded items from Experiment 5, EK produced 3 correct definitions, 18 partially correct definitions and 12 entirely incorrect/no response errors (92% of items). GT produced 0 correct definitions, 14 partially correct definitions and 21 entirely incorrect/no response errors (97% of items). Again, the difference in the number of partially correct definitions was significant ($\chi^2(2) = 5.47, p < 0.05$). These findings are consistent with the suggestion that GT showed rather less effect of set size than EK in degraded word recall because small set sizes improve the recall of partially degraded items to a greater extent than entirely degraded words. Snowden and Neary (2002) reached a similar conclusion in a study that examined the relearning of verbal labels for pictures. They found very poor learning of items that the patients apparently did not comprehend at all, but substantial learning of items that the patients could not name but could demonstrate some knowledge of.

2.8 Serial position effects in the recall of known and degraded words

Experiments 1 to 5 demonstrated, in line with previous studies (Knott et al., 1997, 2000; Patterson et al., 1994), that the likelihood of phonological disintegration in verbal STM is affected by the degree to which an item is semantically degraded. The analysis in this section combines the data from these experiments in order to examine the effect of semantic degradation on the shape of the serial position curve. As noted in Chapter 1, there is considerable controversy about this topic. N. Martin and Saffran (1997) sought to account for the effect of lexical and semantic representations on phonological STM within Dell and O'Seaghda's (1992) interactive activation model of speech production and predicted that semantic factors should have their biggest impact on the early portions of the serial position curve. During ISR tasks, activation within this model spreads up from the phonological layer to lexical and semantic units and then back down again, allowing lexical and semantic activation to improve repetition accuracy. As it takes a number of processing cycles for the activation to spread, the influence of semantics on ISR should be largest for the earliest presented words. Later portions of the serial position curve should be more heavily influenced by phonological factors, which occur more quickly.

There is some empirical support for the claim that semantic factors have the greatest influence in the initial portions of the serial position curve, both from studies of normal performance (Brooks & Watkins, 1990; Watkins & Watkins, 1977) and from neuropsychological studies showing that semantically impaired patients show a reduced primacy effect, whereas phonologically impaired patients show a reduced recency effect (N. Martin & Saffran, 1990; N. Martin & Saffran, 1997; R. C. Martin & Lesch, 1996; Saffran & Martin, 1990). The literature is highly inconsistent however. Hulme et al. (1997) found that the recall difference between high and low frequency words increased towards the *end* of the list, whereas Walker and Hulme (1999) obtained parallel serial position curves for concrete and abstract words. In addition, several neuropsychological studies have found little difference in the shape of the serial position curve for relatively well-known and semantically degraded words (Forde & Humphreys, 2002; Knott et al., 1997). The semantically impaired patients in these previous studies showed primacy effects but negligible recency effects, contrary to the suggestion of N. Martin and Saffran (1997).

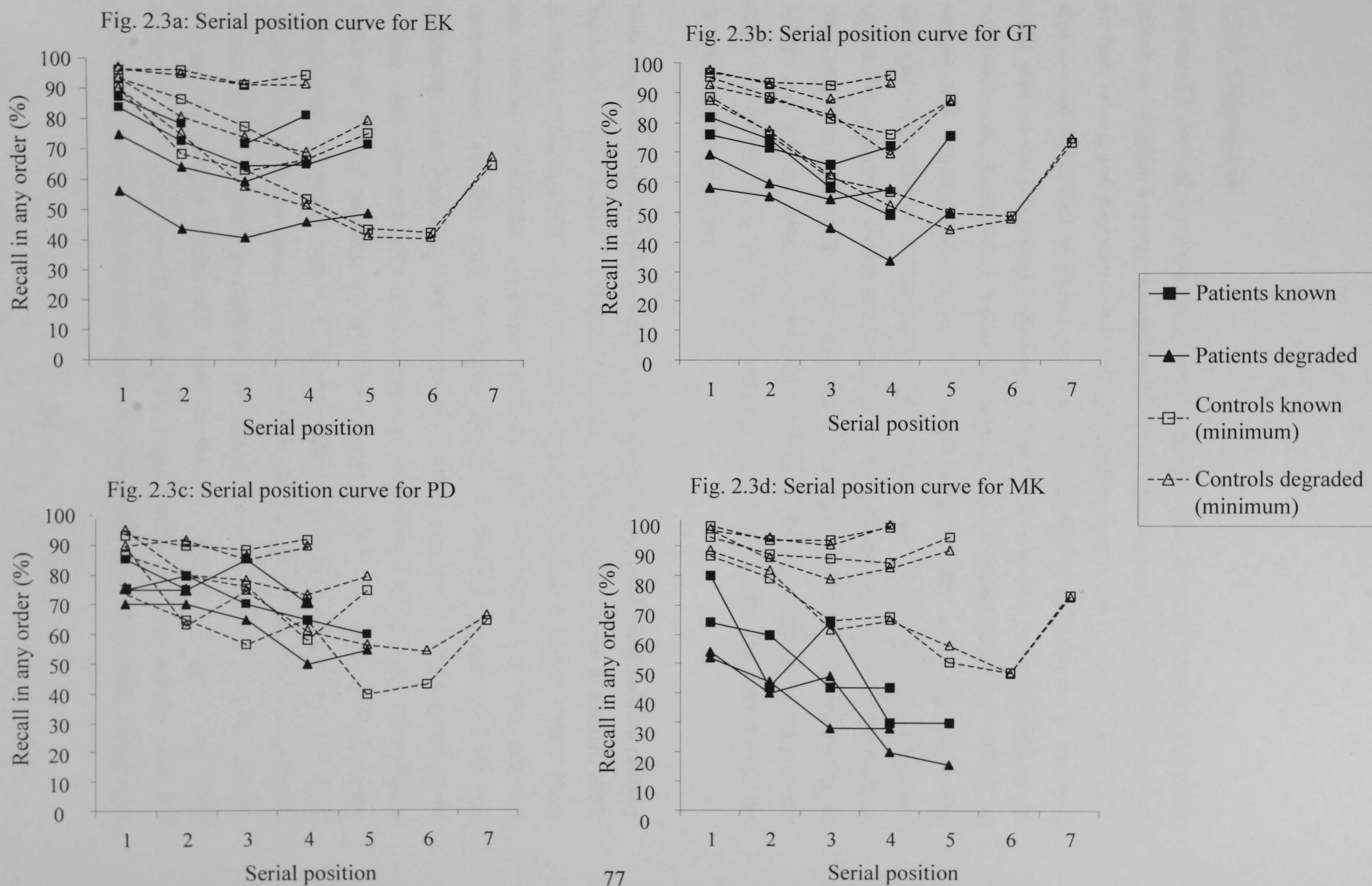
2.8.1 Method

Serial position curves were derived for EK, GT, PD, MK and their controls by amalgamating the data from Experiments 1 to 5 and the results of two similar experiments, not reported here. In every experiment, the patients had recalled matched lists of known and degraded words and these items were examined separately. The largest amount of data was available for EK and GT and the smallest amount of data was available for PD. For EK and GT, there was considerably more data involving four-word lists than five-word lists. For EK and GT, 254 four-item lists and 110 five-item lists were available in the known and degraded conditions. For their controls, 798 four-item lists, 420 five-item lists and 636 seven-item lists were available, combining the data from six different participants. PD was tested on 20 four-item lists and 20 five-item lists. For her controls, 60 lists were available at each list length, combining the data from three participants. MK was tested on 50 four-item lists and 50 five-item lists. For MK's controls, 180 lists were available at each list length, combining the data from six participants.

2.8.2 Results

Figures 2.3a-d show serial position curves for the four patients and their controls. The controls' serial position curves showed the standard pattern of markedly better recall at the beginning and end of the lists. For all four patients, there was no significant interaction between the known-degraded variable and serial position, on both four-item lists (EK, MK and PD: $\chi^2(3) < 1$; GT: $\chi^2(3) = 2.76$, n.s.) and five-item lists (EK, GT and PD: $\chi^2(4) < 1$; MK: $\chi^2(4) = 1.38$, n.s.). EK, GT and PD showed serial position effects that did not differ from their controls for both known words (EK: $\chi^2(4) = 1.34$, n.s.; GT: $\chi^2(4) = 3.72$, n.s.; PD: $\chi^2(4) < 1$) and degraded words (EK: $\chi^2(4) = 1.27$, n.s.; GT: $\chi^2(4) = 2.07$, n.s.; PD: $\chi^2(4) < 1$). These analyses examined recall on five-item lists. In contrast with the other patients, MK did not show normal effects of serial position, for either known words ($\chi^2(4) = 15.99$, $p < 0.01$) or degraded words ($\chi^2(4) = 13.29$, $p < 0.01$). For serial positions 1 to 5, MK's standardised residuals in these analyses were 2.3, -0.6, 1.4, -1.6 and -1.9 for known words and 1.5, 0.5, 1.5, -1.5 and -2.2 for degraded words. These values swing from positive to negative, suggesting that MK's recall was poorer towards the end of the lists; i.e. she did not show the standard recency effect.

Figure 2.3: *Serial position curves for known and degraded words*



2.8.3 Discussion

EK and GT showed relatively typical serial position curves that were broadly parallel for known and degraded words. In contrast, MK's recall fell steadily across serial positions for both known and degraded words and she failed to show a normal recency effect. PD also showed little effect of recency in her recall, although for this patient, the effects of serial position did not diverge significantly from the normal pattern. These findings are consistent with several other studies that have observed diminished recency effects in semantically impaired patients and comparable serial position curves for known and degraded words (Forde & Humphreys, 2002; Knott et al., 1997). Little evidence was obtained to support N. Martin and Saffran's (1997) contention that semantic factors have their biggest impact on the early portions of the serial position curve. According to Martin and Saffran, semantic impairments should reduce the magnitude of the primacy effect but should leave the recency portion of the curve intact. For patient MK, the opposite appears to be true.

How can these observations be reconciled with studies that have found reduced primacy but intact recency effects in semantically impaired patients (N. Martin & Saffran, 1990; N. Martin & Saffran, 1997; R. C. Martin & Lesch, 1996; Saffran & Martin, 1990)? There are several potentially important methodological differences between previous investigations and this study. N. Martin and R. C. Martin's patients generally had substantial phonological as well as semantic impairments, whereas the patients in the present study had relatively intact phonology; they almost never made phonological errors in picture naming or spontaneous speech and had normal digit spans. Consequently, N. Martin and R. C. Martin's patients had much more severe deficits of verbal STM and it was appropriate to test their ISR on very short lists. N. Martin and Saffran (1997), for example, examined the repetition of single items and pairs of items and equated primacy effects with recall of the first word in the pair or the initial phonemes of a single item. In contrast, the patients included in the present study had much higher levels of ISR and were tested on considerably longer lists. This difference in list length may have important consequences for N. Martin and Saffran's (1997) prediction. Even if semantic activation lags behind phonological activation as predicted by Dell and O'Seaghda's (1992) model, over the course of longer lists there may be time for semantic representations to become activated for all the list items, and consequently, semantic deficits may affect every portion of the serial position curve equally.

It is still necessary to explain why the recency effect was diminished in the more severely semantically impaired patients. One possibility is that over the course of an ISR task, the phonological representations of the items to be recalled become noisy and unreliable. In healthy individuals, the appropriate phonological activation can be sustained by lexical and semantic support, as proposed by N. Martin and Saffran (1997). When this form of support is unavailable, however, as in the case of semantically degraded words, accurate phonological representations cannot be maintained for long enough to support recall of items at the end of the list (see Hulme et al., 1997, for a similar suggestion).

2.9 General Discussion

This series of experiments has explored the effect of various methodological factors on the ISR of relatively well-known and semantically degraded words in four patients with SD. Specifically, the research examined the influence of the number of known and degraded words in the lists (set size), the method used to classify items as known and degraded, and the length of the lists to be recalled. In general, the patients recalled the known words more accurately than the degraded words and made a larger number of phonological errors on the degraded words, consistent with the notion that semantic representations play a major role in maintaining the phonological coherence of words in STM. This discussion will examine the methodological variables that did and did not affect the magnitude of the recall difference between known and degraded words and will consider what these findings might suggest about the mechanism underlying the superior recall of known words.

In Experiments 1 and 2, which involved small sets of frequency-matched known and degraded words, two patients, GT and MK, showed a significant advantage for the recall of known words over degraded words, but two other patients, EK and PD, did not. In Experiment 3, the requirement for the known and degraded words to be matched for frequency was relaxed, allowing set size to be substantially increased. Under these conditions, EK did recall significantly more known than degraded words, whereas the controls did not, suggesting that frequency differences per se could not account for the known-degraded difference. Experiments 4 and 5 more directly compared recall in large and small set size conditions and found that when the same words were presented repeatedly, EK's recall improved more substantially for degraded than for known words. EK made fewer phonological errors when set size was small, suggesting that her increased familiarity with the phonological forms of degraded words improved their coherence in STM. Limited set size appeared to account for EK's equivalent recall of known and degraded words in Experiments 1 and 2. As many of the previous studies finding no recall difference between known and degraded words also used small set sizes (see Table 2.1), some of the discrepant results in the literature might be attributable to this factor.

Both EK and GT were included in Experiments 4 and 5, and set size did not affect them in exactly the same way. EK only showed a substantial recall advantage for known over degraded words when set size was large, whereas GT recalled the known words better than the degraded words at every set size. Set size particularly improved EK's recall of degraded words. In contrast, GT's degraded words showed little improvement with small set sizes. These findings point to an effect of semantic knowledge on the degree to which repeatedly presenting the phonological forms of words boosts their immediate recall. Set size might most strongly affect the recall of words that are partially semantically degraded but not completely forgotten. EK's degraded words appear to have fallen into this category, whereas GT's degraded words were more substantially impaired. These results are consistent with those reported by Snowden and Neary (2002). They found that patients with SD are able to re-learn the phonological forms of words that they still partly know but are much less able to re-learn the phonological forms of words that have completely impoverished semantic representations.

How can the effect of set size on the recall of known and degraded words be explained? In healthy participants, recall is enhanced by small set sizes (Coltheart, 1993; Conrad, 1963), presumably because the words in small sets become more predictable as they are repeated. As a result, words from small sets should be easier to encode, retain and recall correctly. Roodenrys and Quinlan (2000) proposed an explanation of the set size effect based on Hulme et al.'s (1997) 'redintegration' theory. According to this account, noisy phonological traces in STM are automatically reconstructed from long-term lexical representations, during the process of recall. The number of lexical candidates in the reconstructive process is reduced when set size is small, increasing the likelihood that the phonological trace will be restored accurately. Similar effects of set size might also be predicted by interactive theories, for example, the interactive-activation account of N. Martin and Saffran (1997). The repeated presentation of items from limited sets might be expected to increase lexical and semantic activation for those items. The model predicts that this enhanced activation will constrain activity at the phonological level, increasing the likelihood that the target is produced correctly.

These theories may also be able to account for the differential effect of set size on known and degraded words. Words that are severely degraded do not benefit from semantic support in ISR tasks, and consequently their phonological elements are not produced in the correct configuration. Set size would be expected to have little impact on these words with completely degraded semantics, as it is the lexical-semantic representations themselves that are thought to underpin the set size effect. Set size might also be expected to have little impact on the recall of very well known words, as these items will be adequately supported by intact semantic representations. Partially degraded words would be expected to derive the most support from small set sizes, as repeated presentation of these words might boost any residual lexical-semantic activation that still plays a role in maintaining phonological coherence in STM. Roodenrys and Quinlan (2000) found that set size only affected the recall of lower frequency items in normal participants, presumably because higher frequency items were adequately supported by their more accessible lexical representations. This result may be analogous to the effect of set size on the recall of well-known and partially degraded words in patients with SD.

We can also consider the impact of the methods used to select items as known and degraded. Experiments 1 and 2 employed rather different semantic tests to classify words as known and degraded, but in both experiments, GT and MK recalled the known words at a substantially higher level than the degraded words, whereas EK and PD did not. This consistency suggests that the methods used to select known and degraded items may not have a major impact on the magnitude of the known-degraded recall difference in terms of accuracy. Some clear differences did emerge, however, between the two sets of words. For the naming and definitions words, phonological errors occurred more frequently for the degraded than the known items, even for EK and PD who did not show an accuracy difference. There was no such error difference for the words selected according to synonym judgements, even for patients who showed an accuracy difference. The synonym judgement words were recalled more poorly than the naming and definitions words, were more strongly affected by set size and were characterised by frequent phonological errors affecting both known and degraded items. These findings suggest that the meaning of the synonym judgement words may have been generally more degraded than the naming and definitions words.

Experiments 1 and 2 also examined the impact of list length on the recall of known and degraded words. The size of the known-degraded recall difference did not vary systematically with list length and this factor had relatively little effect on the occurrence of phonological errors. In contrast, the number of non-phonological errors (predominantly omissions) rose steadily as list length was increased. There was no evidence to suggest that the coherence of items in STM was diminished in conditions of high phonological load. Phonological coherence was affected primarily by the type of material to be retained, i.e. the status of the words as known or degraded, and not by the amount of material. This finding is consistent with the suggestion that although semantic degradation impairs ISR performance in SD, the underlying phonological STM mechanism is intact, allowing these patients to show normal effects of phonological similarity and word length in ISR (see Knott et al., 2000, and Chapter 4). The patients were able to retain phonological representations of a relatively normal number of items in STM, although the coherence of these representations was weakened for degraded words, allowing phonological elements to migrate between the list items.

This work has demonstrated that methodological factors, such as set size, can affect the magnitude of known-degraded recall differences in SD patients. In fact, almost all of the previous failures to find such differences can be accounted for by the use of small set sizes (see Table 2.1). The one exception is McCarthy and Warrington's (2001) patient MNA, who failed to show a recall difference between known and degraded words even though there were 30 items in each set. Like patient PD, MNA had relatively normal recall accuracy for words that she understood poorly. It is important to note, however, that both PD and MNA made frequent phonological errors in ISR, suggesting that their semantic degradation did affect the phonological coherence of items in STM.

Cases such as PD and MNA, who have good immediate recall despite severe semantic deficits, suggest that individual differences in dimensions other than semantic degradation may contribute to the degree of ISR impairment in this condition. One potentially important factor is phonology: GT and MK, who showed the largest known-degraded recall differences, may have had phonological deficits in addition to their semantic impairments. The recall of degraded words would be particularly sensitive to phonological impairment, as these words would derive little support from semantics. Several authors have argued that additional phonological or lexical impairments are required to produce phonological breakdown in ISR for degraded words (Knott et al., 1997; McCarthy & Warrington, 2001), challenging Patterson et al.'s (1994) claim that semantics plays a necessary and major role in constraining phonological activation in STM. However, if the patients in this study had any phonological deficits, they must have been relatively subtle. The patients had normal digit spans (see Chapter 3) and very rarely made phonological errors in spontaneous speech or picture naming. In addition, they showed effects of phonological similarity in ISR that were within the normal range (see

Chapter 3), suggesting that their verbal STM relied heavily on a phonological code as in healthy participants (it was only possible to test EK, GT and MK in this experiment). Although the patients made an abnormal number of phonological errors in their recall of known words, these errors may have reflected partial semantic degradation rather than additional phonological deficits. If a continuum of semantic degradation underlies the known/degraded distinction, comprehension of known words should not be perfect and phonological errors should occur on these items. On the other hand, the lack of a recency effect in MK's ISR might be taken as evidence for a phonological deficit, as this pattern is associated with phonological impairments in aphasic populations (N. Martin & Saffran, 1997). We will return to this important issue of phonology in patients with SD in Chapter 4.

While it remains possible that SD patients who show substantial known-degraded recall differences have additional subtle phonological deficits, it is also possible that individuals who fail to show an ISR impairment for degraded words, like PD and MNA, have exceptional phonological abilities that can maintain the phonological elements of words in the correct configuration with minimal support from semantics. Indeed, both PD and MNA had exceptional digit spans of around eight items and PD actually outperformed her controls on a digit span task (see Chapter 3), suggesting that these patients did have superior phonological abilities. An analogy can be drawn from the domain of reading, where it has been argued that semantics plays a major role in translating between orthography and phonology (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996). According to this account, semantically impaired patients should be poor at reading low frequency words with atypical spelling-to-sound correspondences, and indeed many semantically impaired patients do show this pattern (Graham, Hodges, & Patterson, 1994; Patterson & Hodges, 1992). There are, however, a few semantically impaired cases who are normal at reading low frequency exception words (Cipolotti & Warrington, 1995; Lambon Ralph, Ellis, & Franklin, 1995). According to Plaut (1997), these cases had highly developed phonological pathways that mastered the problem of learning to read without the usual level of semantic involvement. Similarly, PD and MNA may have had good phonological skills as their verbal STM abilities developed, causing their mature phonological systems to be less reliant on semantic support. Without independent evidence of excellent phonological processing in patients like PD and MNA, however, this hypothesis must remain conjecture. In Chapter 4, multiple measures of phonological processing ability are presented for six SD patients, including data for EK and GT collected approximately a year after the testing presented here was carried out. Unfortunately, it was not possible to include PD in this study because she did not wish to participate in further research.

In summary, this chapter has explored the impact of methodological factors on the size of the known-degraded recall difference in four patients with SD, in order to investigate possible reasons for the discrepancies in the results of previous studies. It appears that many of the failures to find ISR differences between known and degraded words could be accounted for by methodological factors; in particular, set size.

3 A category specific advantage for numbers in verbal short-term memory

3.1 Introduction

The work presented in Chapter 2 explored the effect of semantic knowledge on immediate serial recall (ISR). It was found, in line with previous studies (e.g., Knott, Patterson, & Hodges, 1997; Knott, Patterson, & Hodges, 2000; Patterson, Graham, & Hodges, 1994) that patients with semantic dementia (SD) make a larger number of phonological errors in ISR for words that are semantically degraded. This finding supports the notion that long-term semantic representations play a role in maintaining the phonological coherence of words in STM.

It is striking that digit span is typically unimpaired in SD whereas word span is severely impaired. Warrington (1975) described two patients with word spans of four but digit spans of nine and seven. Similarly, patient AM (Knott et al., 1997) recalled lists of five digits almost perfectly but only a quarter of lists of the same length when they were composed of letters or high frequency words. Digit span also remains relatively stable in the face of marked semantic decline (Knott et al., 2000).

This difference between digit span and word span could arise for a number of reasons. One intriguing possibility is that ISR for digits is relatively preserved because digits are understood well in comparison with other categories of word. That is, the difference might be equivalent to the ISR difference between well-known and semantically degraded words examined in the last chapter. In line with this suggestion, some recent studies have found that number knowledge is relatively spared in semantic dementia (Butterworth, Cappelletti, & Kopelman, 2001; Cappelletti, Butterworth, & Kopelman, 2001; Cappelletti, Kopelman, & Butterworth, 2002; Crutch & Warrington, 2002; Diesfeldt, 1993). Cappelletti et al. described patient IH, who displayed well-preserved numerical understanding despite severe semantic impairments in other domains (Butterworth et al., 2001; Cappelletti et al., 2001; Cappelletti et al., 2002; Crutch & Warrington, 2002; Diesfeldt, 1993). He was able to read, write and transcode Arabic numerals and English number words, and make accurate judgements about numerical magnitude. His calculation abilities also were relatively intact although he made some errors on multiplication and division problems, particularly those involving larger numbers. In contrast, he was markedly impaired at answering number fact questions that entailed knowledge of non-numerical entities (e.g., how many months are in a year?). Cappelletti et al. argued that IH had intact representation of number, but degradation of number facts and procedural knowledge of calculation. Crutch and Warrington (2002) observed a strikingly similar pattern of abilities and deficits in a second patient with SD.

The cortical atrophy in SD predominantly affects the anterior and inferior temporal lobes bilaterally, and the temporal poles in particular (Galton et al., 2001; Mummery et al., 2000). In contrast, functional imaging and neuropsychological studies suggest that knowledge of numbers is associated with a network of parietal areas, in particular the inferior intraparietal area (Dehaene & Cohen, 1995; Dehaene, Dehaene-Lambertz, & Cohen, 1998). Damage to the intraparietal area is associated with acalculia (e.g., Cipolotti, Butterworth, & Denes, 1991; Dehaene & Cohen, 1997; Delazer & Benke, 1997; Takarama, Sugishita, Akiguchi, & Kimura, 1994; Warrington, 1982), which has been characterised as an impairment of numerical representations, rather than calculation difficulties per se (Dehaene & Cohen, 1997). In functional imaging studies, number and calculation tasks produce activity in parietal regions including the intraparietal area (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Dehaene et al., 1996; Stanescu-Cosson et al., 2000). The amount of intraparietal activation in numerical comparison tasks is closely correlated with numerical distance, suggesting that this activation corresponds to a conceptual representation of numerical magnitude (Pinel, Dehaene, Rivière, & LeBihan, 2001). Collectively, these neuropsychological and neuroimaging findings indicate that some aspects of number representation may be largely independent of the temporal lobe semantic system that becomes degraded in SD.

The relative preservation of number knowledge in SD is not the only possible cause of the difference between digit and word span. Digits are higher in frequency than the words typically used to assess span and ISR performance in SD is strongly affected by frequency (Knott et al., 1997; Knott et al., 2000; McCarthy & Warrington, 2001). In addition, SD patients show effects of imageability in ISR (Knott et al., 1997) and the words and digits used in span tests are not typically matched for imageability. Thirdly, word span is normally tested with non-repeating items whereas digit span uses a restricted set of nine items. Smaller set sizes improve recall in SD patients as well as normal participants (Knott et al., 1997), so set size may contribute to the better ISR for digits. In addition, digits are drawn from a closed semantic set where as the words used in span tasks can be drawn from many semantic categories. A fifth potential factor is that numbers form an ordered sequence whereas words do not. Finally, normal subjects show better recall of digits than words (Brener, 1940) and it is not clear from the existing reports whether the difference between digit and word ISR in SD patients is more substantial than in controls.

This chapter examined ISR of number and non-number words matched for frequency, imageability, word length, set size, and size of semantic category in four SD patients, and made a direct comparison between patients and controls. The materials were extended the to determine whether better performance with single-digit numbers generalised to lower frequency multi-digit numbers. Comprehension of the number and matched non-number words was assessed in order to examine whether a category specific difference in comprehension could underlie the ISR results. Case descriptions for the four patients are provided in section 2.2.

3.2 Experiment 1: Immediate serial recall of digits and matched words

This experiment aimed to determine whether the ISR difference between single-digit numbers and non-number words would persist when the items were matched for length, frequency, imageability and set size.

3.2.1 Method

Nine words were matched on an item-by-item basis to the digits 1-9 for syllable length and word frequency using lemma counts from the Celex database (Baayen, Piepenbrock, & van Rijn, 1993). Many of these high frequency words were abstract in nature, so a second set of words was selected to match for frequency, imageability and syllable length using imageability counts from the MRC psycholinguistic database (Coltheart, 1981). These two sets of words are reproduced in Appendix 5.

A variety of list lengths were tested. There were ten lists at each length. The length of the lists depended on each patient's ISR abilities. EK was tested on four to seven items. GT was tested on four to eight items. PD had an exceptional digit span and was tested on six to eight items (although for the second set of words, eight item lists were not tested due to time constraints). MK performed more poorly on ISR tasks, so was tested on four to six items. The numbers and words were yoked so that matched items appeared in the same position in each list. The single-digit numbers and frequency-matched words were presented in blocks using an ABBA design. The frequency and imageability-matched words were tested on a separate occasion. Testing was extended to include shorter list lengths for words and longer list lengths for digits, allowing the patients' errors just beyond span to be compared across the materials (although due to time constraints, this process was not completed for PD). EK was tested on lists of eight numbers. GT was tested on nine numbers and two to three words. PD was tested on nine numbers and five words from the second set. MK was tested on seven to eight numbers and three words. The items were read aloud at a rate of one word per second for immediate serial recall.

Twelve healthy control participants were also tested on these materials. Three control participants were matched to each of the four patients on age and years of education. They were tested on list lengths from four to nine items for all three types of material. The materials were presented in blocks using an ABCCBA design.

3.2.2 Results

3.2.2.1 Recall accuracy

Figure 3.1 shows the percentage of single-digit numbers and non-number words recalled in the correct order by patients and controls. Table 3.1 indicates span for all the materials tested in this series of experiments, and shows that the pattern of results was similar whether performance was measured in terms of list or item accuracy.

Table 3.1: *Spans for the number and non-number words used in Experiments 1 – 3*

	EK	GT	PD	MK	Control mean (range)
Digits (Exp. 1)	6	7	8	6	6.08 (5 – 8)
High freq words 1 (Exp. 1)	4	2*	$\leq 5^{\dagger}$	3*	4.75 (4 – 7)
High freq words 2 (Exp. 1)	4	3*	5	3*	4.67 (4 – 6)
Low freq numbers (Exp. 2)	3	3	-	2*	3.67 (3 – 4)
Low freq words 1 (Exp. 2)	2*	2*	-	1*	4.08 (3 – 5)
Low freq words 2 (Exp. 2)	3	3	-	2*	4.33 (3 – 5)
Numbers (Exp. 3)	4	4	-	3	3.83 (3 – 5)
Face-parts (Exp. 3)	3*	3*	-	1*	4.75 (4 – 6)

* denotes abnormal performance. Span was defined as the length at which at least half the lists were repeated correctly.

† PD was above span on 6 items, but shorter lengths were not tested.

Words 1 = frequency matched. Words 2 = frequency and imageability matched (Exp. 1) or frequency matched and high imageability (Exp. 2).

Figures 3.1a-d: *Recall of single-digit numbers and matched words by patients and controls in Experiment 1*

Fig. 3.1a. EK's recall of digits and words

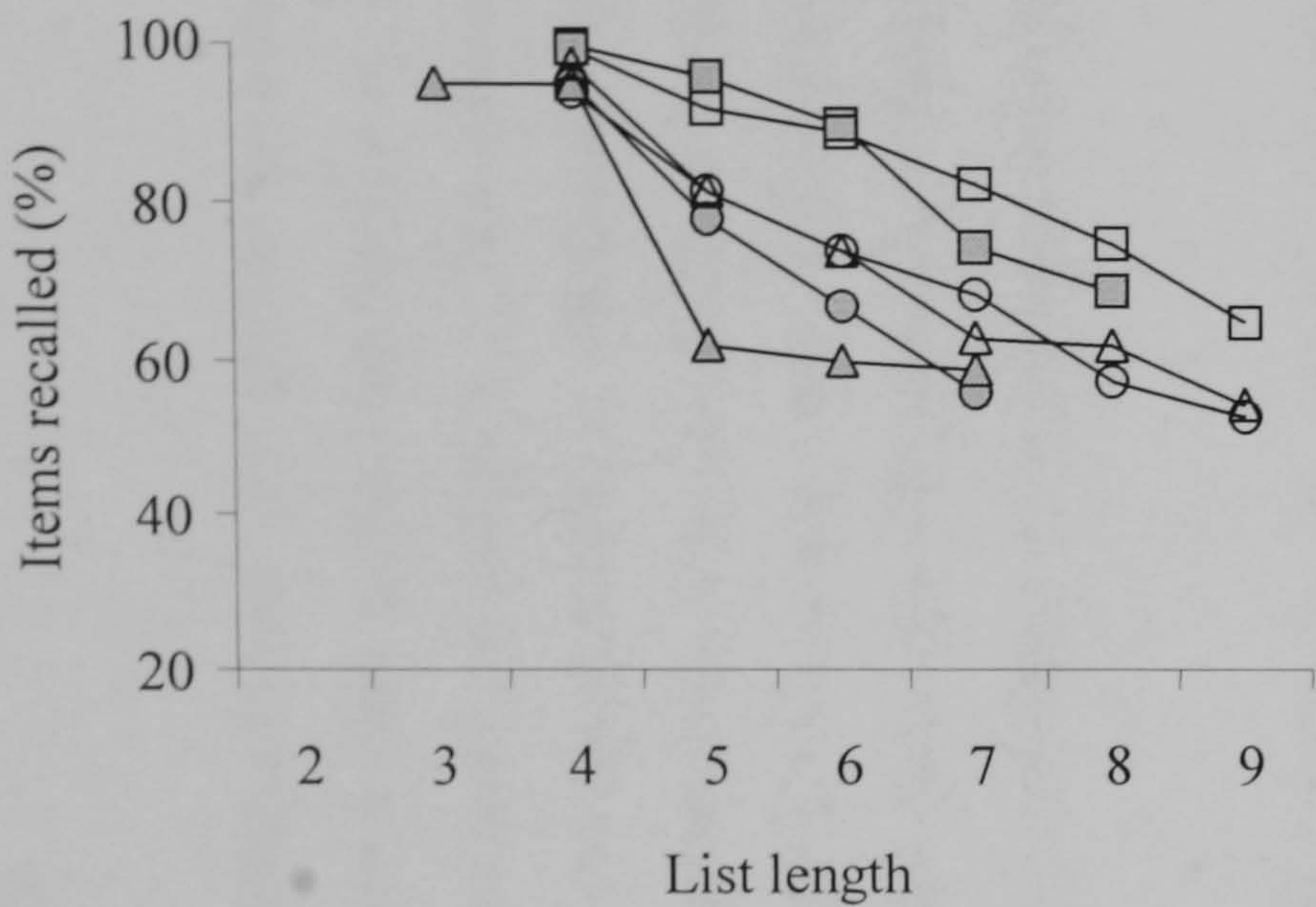
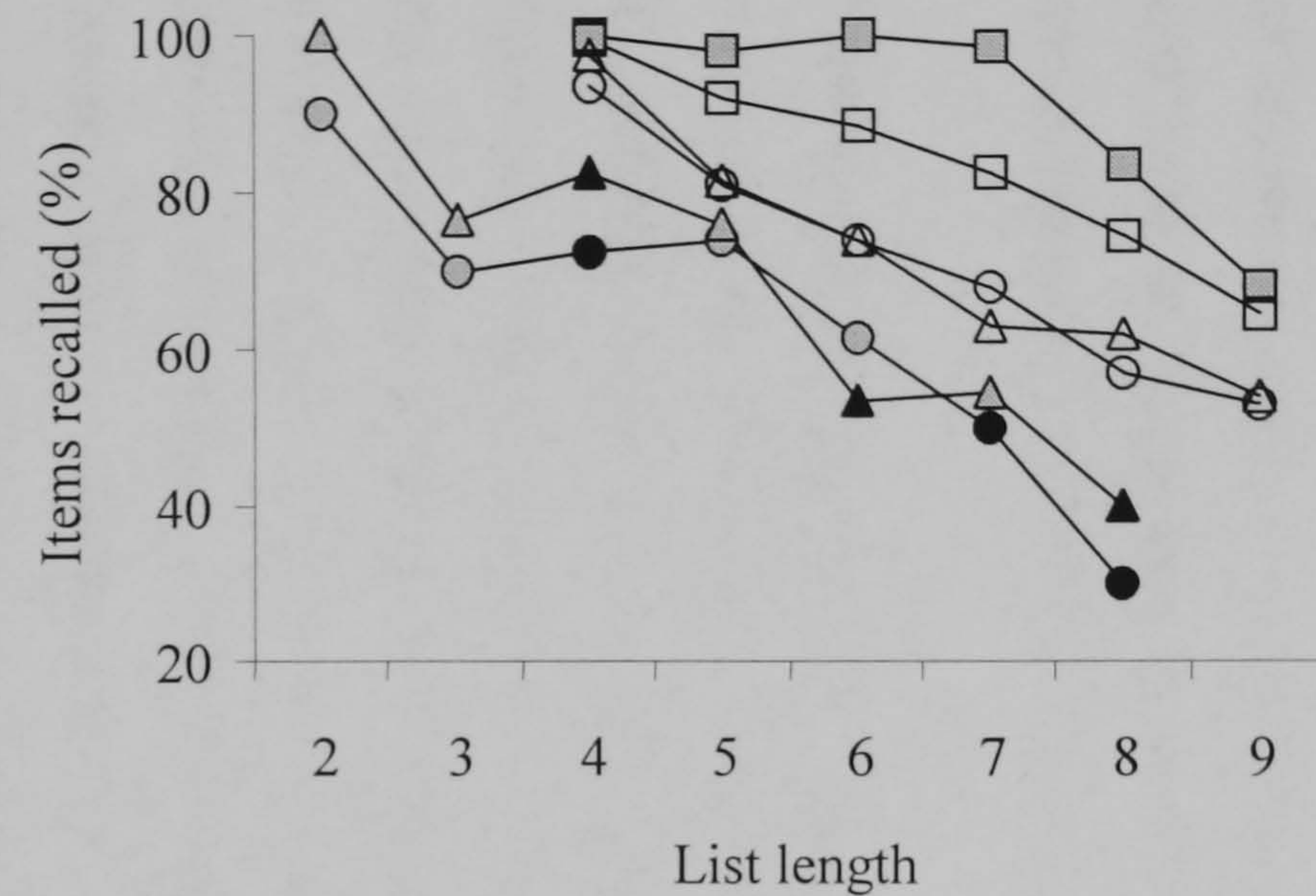


Fig. 3.1b. GT's recall of digits and words



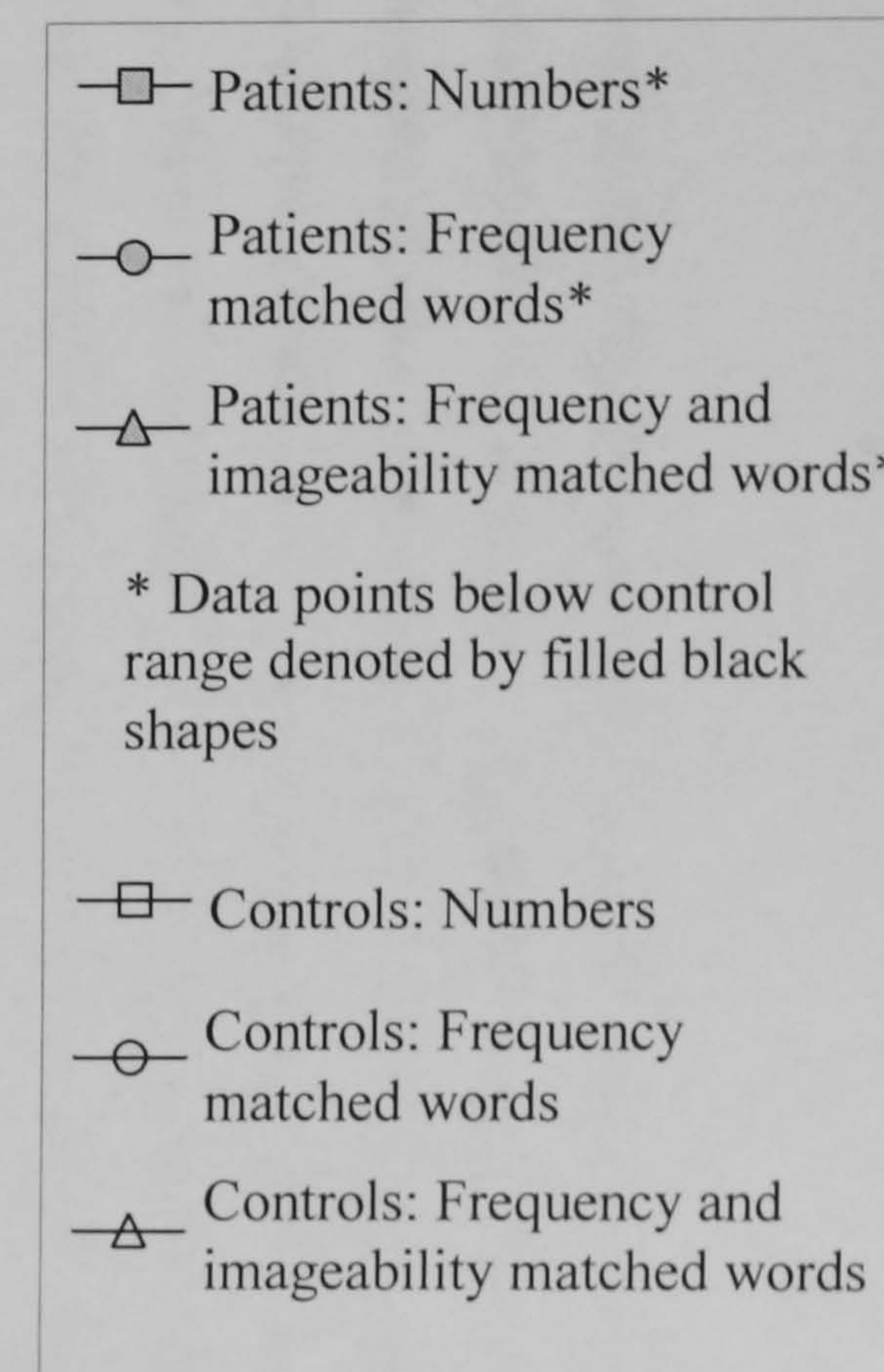


Fig. 3.1c. PD's recall of digits and words

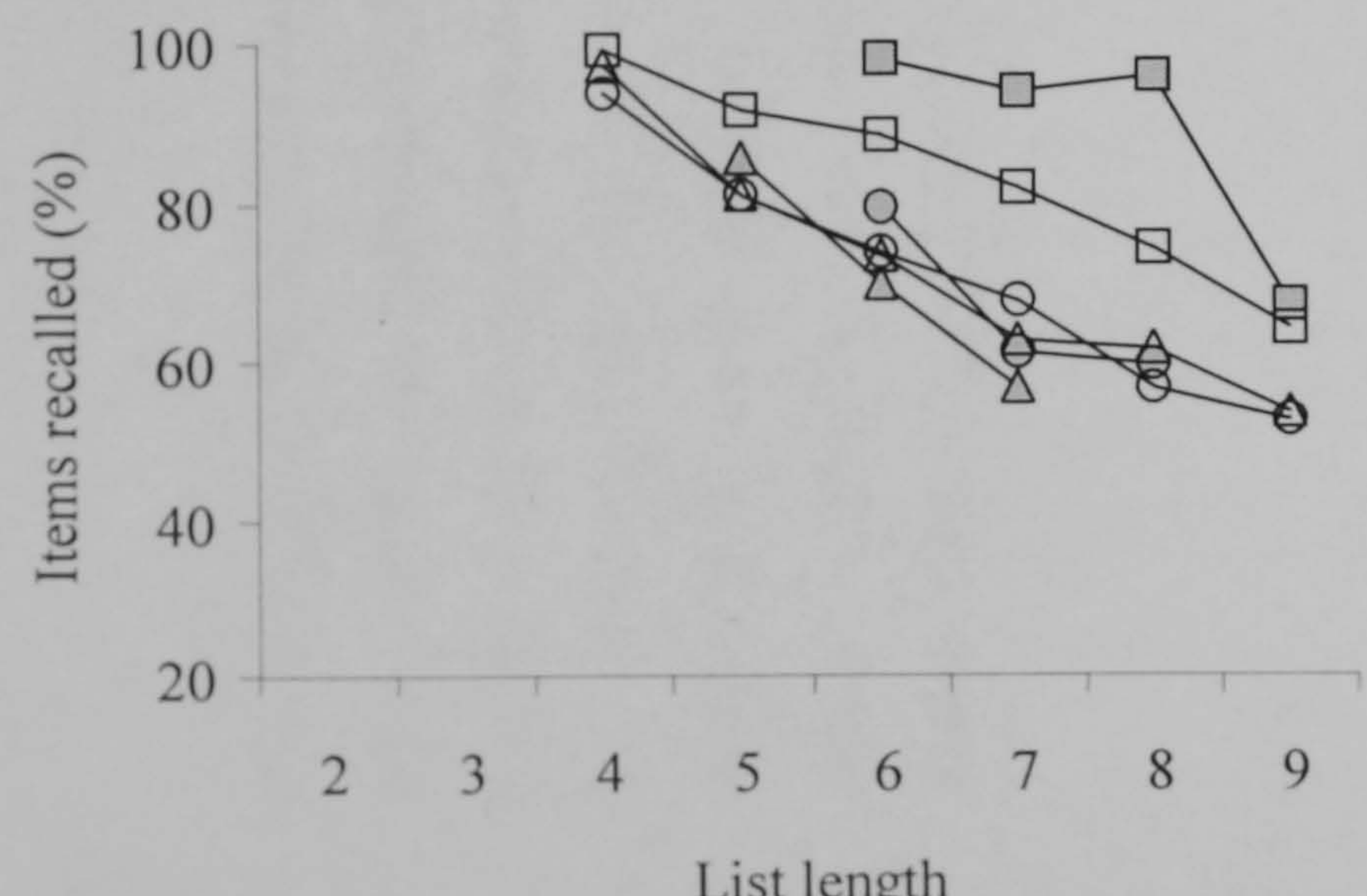
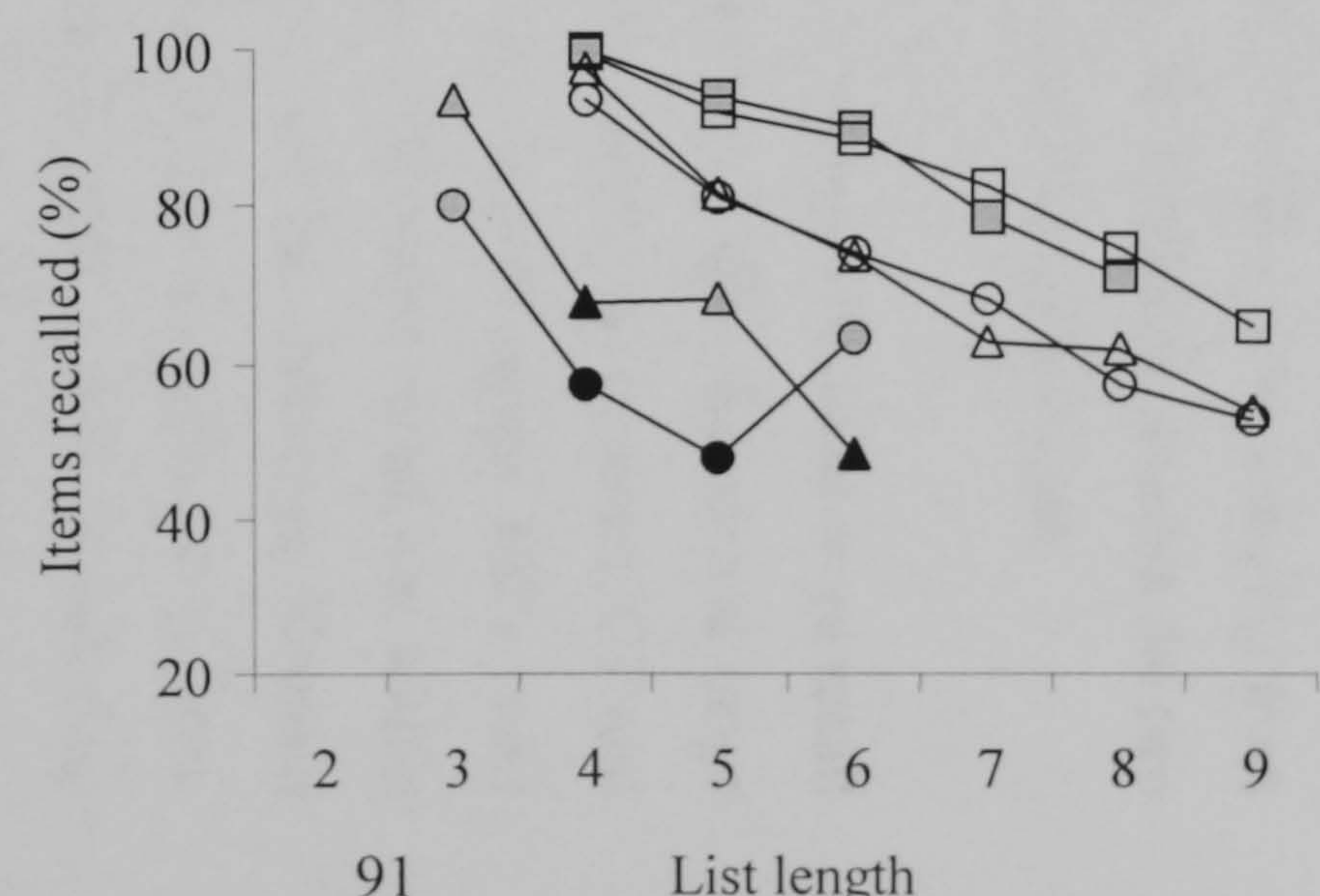


Fig. 3.1d. MK's recall of digits and words



The controls recalled the single-digit numbers better than the non-number words, even though they were matched for frequency, imageability and set size. The advantage for numbers was reliable relative to both the first set of frequency-matched words ($t(11) = 7.63, p < 0.001$) and the second set of frequency and imageability-matched words ($t(11) = 9.10, p < 0.001$). There was no difference in recall between the two sets of non-number words ($t(11) < 1$). The patients also showed better recall of the numbers than the words. For all four patients, statistical contrasts between the recall of single-digit numbers and both sets of non-number words were significant at $p < 0.001$, with t values ranging from 3.43 to 10.65. None of the patients showed a significant recall difference between the two sets of non-number words (all $t < 1$). These analyses collapsed across list length.

Figure 3.1 shows that the patients had normal single-digit recall but impaired word recall relative to the controls. The recall of single-digit numbers was within the normal range for EK, GT and MK, while PD actually outperformed the controls. In contrast, GT and MK had markedly impaired word recall that fell below the normal range. EK's word recall scores were right at the bottom of the normal range. PD's word recall was within the normal range, but she had an exceptional digit span, and the *difference* between numbers and non-number words fell just outside the normal range on some list lengths. The maximum difference between the percentage of single-digit numbers and non-number words recalled by the control participants was 36%. The largest difference was 34% for EK, 54% for GT, 37% for PD and 46% for MK.

3.2.2.2 Errors committed on digits and words

If the difference between digit and word span corresponds to an ISR difference between known and degraded items, fewer phonological errors should occur in the recall of single-digit numbers, as the more robust long-term representations of these items should help to hold their phonology in place (Patterson et al, 1994). There was, however, no straightforward way to compare the patients' digit and word errors because, except for very long list lengths, the patients were at ceiling on single-digit recall. On lengths where the patients made enough digit errors to analyse, their performance on the word items was abysmal. If digit and word errors were compared at a fixed length, observed differences could result from the discrepancy in difficulty. Therefore, errors on single-digit numbers and non-number words were compared on list lengths that were just above span for the two types of material.

Incorrect responses were categorised as omission, order, repetition, intrusion, phonological and unrelated errors. Omission errors were calculated by subtracting correct responses and other error types from the number of items presented. Responses were counted as order errors if the word produced was a target word occurring somewhere else in the sequence. Repetition errors were target words produced more than once. Intrusion errors were items presented in a previous list. A phonological error reproduced at least 50% of the phonemes from the target word (e.g., 'bread' → 'bed', 'sorry' → 'forry'). Unrelated errors did not fall into any of these categories. Errors of this type were most commonly patient responses that did not reach the criteria for a phonological error (e.g., 'council' → 'cathert').

Table 3.2 indicates the total number of errors of each type, for each individual patient and for the controls as a group, across three list lengths: span, and one and two items beyond span. Span was defined as the longest length at which at least 5/10 lists were repeated correctly. PD was not tested on all the lengths necessary for this method. She was not tested on single-digit numbers at a length of two items beyond span, or on lengths short enough to obtain span for the first set of words. PD's digit scores are an amalgamation of span and one length beyond span, and her scores on the first set of words are a combination of the two shortest lengths tested. Fewer data contributed to PD's scores, so they were scaled up (multiplied by 1.5) to allow a rough comparison to be made with the other patients.

Table 3.2: *Errors on single-digit numbers and matched words in Experiment 1*

	EK			GT			PD			MK			Control mean (range/max)		
Condition	Dig	W1	W2	Dig	W1	W2	Dig†	W1†	W2	Dig	W1	W2	Dig	W1	W2
Span	6	4	4	7	2*	3*	8	≤5	5	6	3*	3*	6.1 (5-8)	4.8 (4-7)	4.7 (4-6)
No. items	210	150	150	240	90	120	170	110	180	210	120	120	212.5 (180-270)	172.5 (150-240)	170.0 (150-210)
Phonological	0	5	6*	0	18*	21*	0	21*	7*	0	28*	22*	0	2.9 (13)	0.8 (4)
Unrelated	0	1	1	0	2	2	0	0	5	0	9*	8*	0	0.3 (2)	0.9 (5)
Omission	9	13	14	14	1	0	7.5	18	26*	18	5	2	11.3 (35)	13.3 (40)	13.4 (25)
Order	27	12	20	15	0	2	25.5	6	20	13	0	0	14.8 (29)	10.9 (25)	10.3 (20)
Repetition	10	0	2	13	0	1	22.5*	10.5	6	6	3	1	7.6 (20)	5.8 (13)	4.7 (16)
Intrusion	3	4	2	2	1	1	0	6	1	8	4	0	3.8 (10)	8.4 (18)	10.2 (22)

* denotes abnormal performance. Figures refer to the total number of errors.

Dig = single-digit numbers. W1 = frequency-matched words. W2 = frequency and imageability-matched words.

† Fewer data contributed to PD's scores, so they were scaled up (multiplied by 1.5) to allow a rough comparison with the other patients and controls.

The error rates for all four patients were either within the normal range or nearly normal for the recall of single-digit numbers. The majority of errors were omissions, order errors, repetitions and intrusions, for both patients and controls. There was virtually a complete absence of phonological and unrelated errors for both groups. The controls' errors to non-number words followed a similar pattern, although there were more intrusions (word set 1: $t(11) = 3.55$, $p < 0.01$; word set 2: $t(11) = 4.09$, $p < 0.01$) and more unrelated errors (word set 1: $t(11) = 2.91$, $p < 0.05$; word set 2: $t(11) = 2.42$, $p < 0.05$) in word recall.

The four patients showed a pattern of errors in their recall of non-number words that was very different both from controls' word recall and from their own number recall. For GT, PD and MK, the number of phonological errors exceeded the control range on both word sets. For EK, the number of phonological errors was outside the control range on the second word set. MK also made a large number of unrelated errors that failed to reach the criterion for a phonological error but may have occurred for similar reasons. In contrast, the numbers of omission, order, repetition and intrusion errors were within the normal range.

For all four patients, there were reliable differences between the pattern of errors in single-digit and non-number word recall. Out of eight possible contrasts between a patient's pattern of errors in digit vs. word recall (digits vs. word set 1 and digits vs. word set 2 for each of the four patients), all eight revealed a statistically reliable difference at $p < 0.01$ or less, with chi-squared values ranging from 14.5 to 74.6. Furthermore, the standardised residuals for phonological errors on the non-number words were high in all cases, suggesting that this error category was a major contributor to the significant chi-squared values.

3.2.3 Discussion

The four SD patients showed poor word recall relative to their excellent performance with digits, even when the materials were matched for frequency, imageability, word length and set size. In addition, phonological errors occurred frequently in the recall of non-number words but extremely rarely in the recall of single-digit numbers. These differences are reminiscent of those reported previously for known and degraded words (Knott et al., 1997; Patterson et al., 1994. and see Chapter 2).

The results of this experiment leave us with a puzzle – why are SD patients normal at repeating sequences of single-digit number words but impaired at repeating non-number words if this difference cannot be accounted for by frequency, imageability, word length or the number of items in the set? The following experiments investigated other possible reasons for the difference, in particular the idea that SD patients understand number words better than they do non-number words. Experiment 2 examined the ISR for lower frequency multi-digit number words, like billion and ninety, together with matched non-number words. Multi-digit numbers are expected to be recalled and comprehended more poorly than single-digit numbers because 1) low frequency words and concepts typically degrade earlier in the course of SD (Funnell, 1995) and 2) multi-digit numbers refer to more difficult numerical concepts (Dehaene & Mehler, 1992). However, lower frequency non-number words should also engender poor recall if semantics makes a major contribution to phonological coherence, and hence the difference between the material types might remain.

This experiment also addressed one concern about the interpretation of the previous study. In Experiment 1, both the healthy controls and the SD patients showed better recall of single-digit numbers than non-number words: therefore, it is possible that the patients' specific ISR impairment for non-number words occurred simply because this task was harder. In the following experiment, the use of lower frequency multi-digit numbers and matched words circumvented this problem, because the normal recall advantage for number words was eliminated.

3.3 Experiment 2: Immediate serial recall of low frequency numbers and matched words

This experiment compared the recall of low frequency multi-digit numbers and matched non-number words in order to determine if the superior recall of single-digit numbers would extend to this new material.

3.3.1 Method

The nine lowest frequency number words in English that were whole words rather than compounds of words (e.g. thirteen, not thirty seven) were selected using lemma counts from the Celex database (Baayen et al., 1993). The numbers were compared with words matched on an item-by-item basis for syllable length and frequency. Imageability ratings were not available for many of these words (using the MRC corpus, Coltheart, 1981), so a second set of frequency-matched, high imageability words was selected. The items are reproduced in Appendix 6.

As in Experiment 1, the lists of numbers and words were yoked so that matched items appeared in the same position within a list. There were ten lists at each length. EK and GT were tested on list lengths from two to seven items, and MK was tested on two to six items (as her performance was poorer). In every case, the numbers and both sets of words were tested at each list length. MK was additionally tested on a single item from the first set of non-number words, to determine her span in this condition. PD was not available to participate in this or subsequent experiments. The numbers and frequency-matched words were tested in blocks using an ABBA design. The frequency-matched, high imageability words were tested separately. The 12 control participants described for Experiment 1 were also tested on these materials, using list lengths from three to eight items arranged in an ABCCBA design.

3.3.2 Results

3.3.2.1 Recall accuracy

Figure 3.2 indicates the number of items recalled in the correct order by patients and controls. In contrast to Experiment 1, the control participants did not show an advantage for repeating number words over non-number words. In fact, the controls showed a highly significant advantage for repeating the non-number words over the low frequency multi-digit numbers (collapsing across list length: numbers vs. word set 1: $t(11) = 6.55$, $p < 0.001$; numbers vs. word set 2: $t(11) = 8.74$, $p < 0.001$). In addition, the controls recalled the higher imageability words (set 2) more accurately than the lower imageability words (set 1: $t(11) = 5.51$, $p < 0.001$).

EK showed an ISR advantage for the multi-digit numbers over the first set of non-number words which approached significance (collapsing across list length: $t(114) = 1.94$, $p < 0.06$). She showed no difference between the numbers and the second set of non-number words ($t(113) < 1$). GT's ISR performance was significantly better for the numbers than the first set of words ($t(108) = 3.17$, $p < 0.01$) and he showed no difference between the numbers and the second set of words ($t(112) = 1.53$, n.s.). MK recalled the multi-digit numbers better than the words from both set 1 ($t(98) = 5.07$, $p < 0.0001$) and set 2 ($t(98) = 2.22$, $p < 0.05$). Therefore, all three patients recalled the numbers as well as or better than the non-number words, whereas the controls recalled the non-number words more accurately than the numbers.

Figures 3.2a-c: *Recall of low frequency multi-digit numbers and matched words by patients and controls in Experiment 2*

Fig. 3.2a. EK's recall of low frequency numbers and words

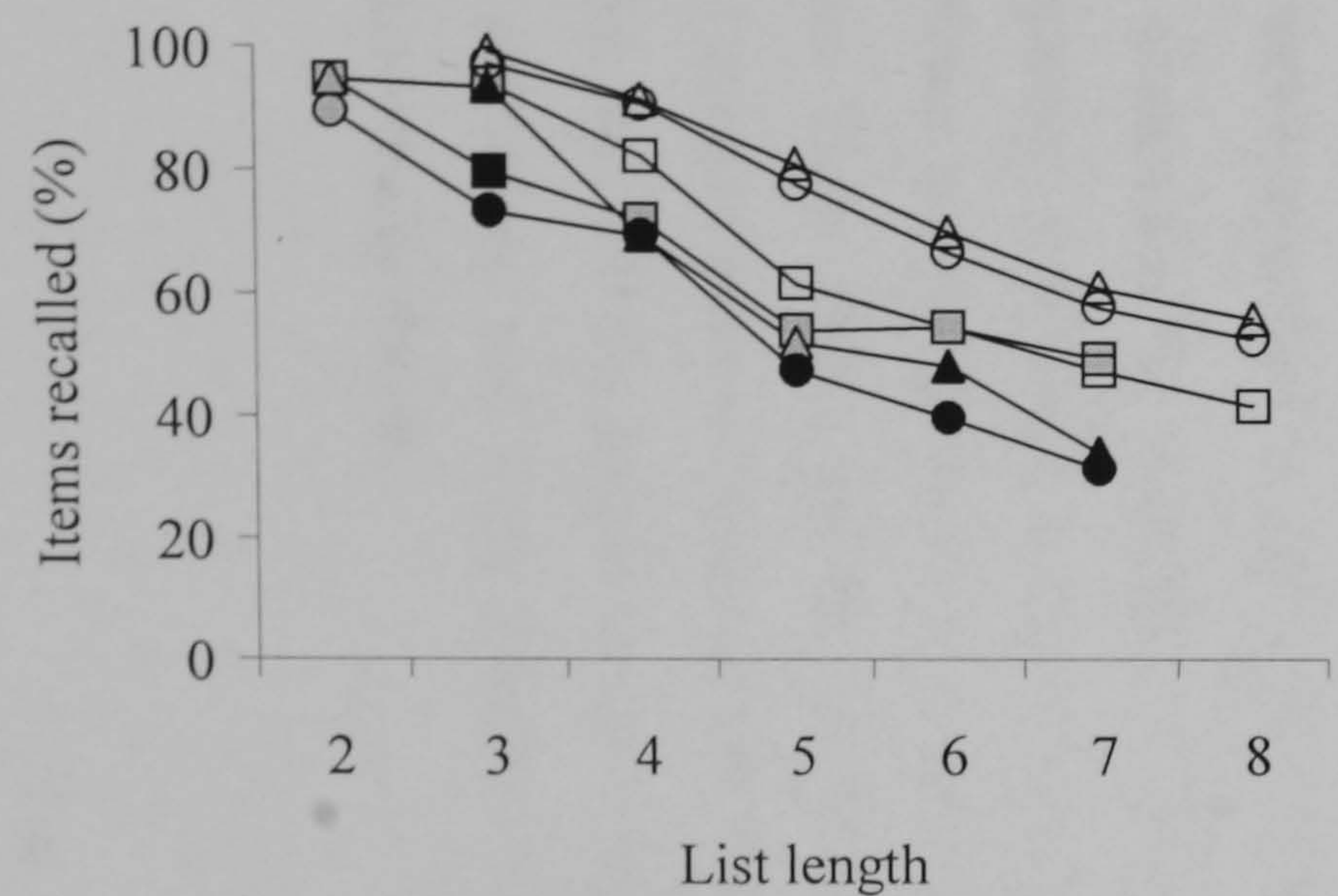
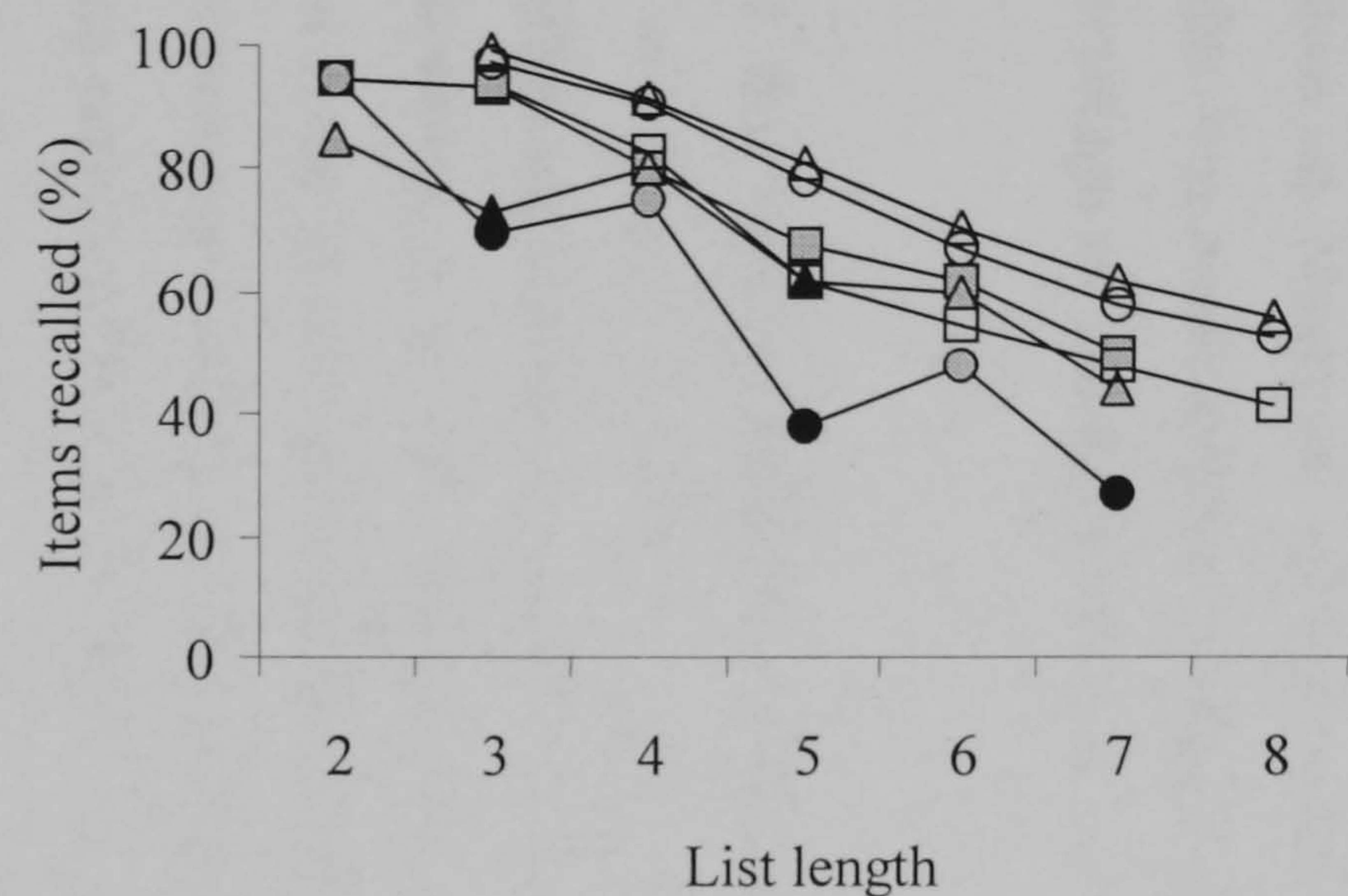


Fig. 3.2b. GT's recall of low frequency numbers and words



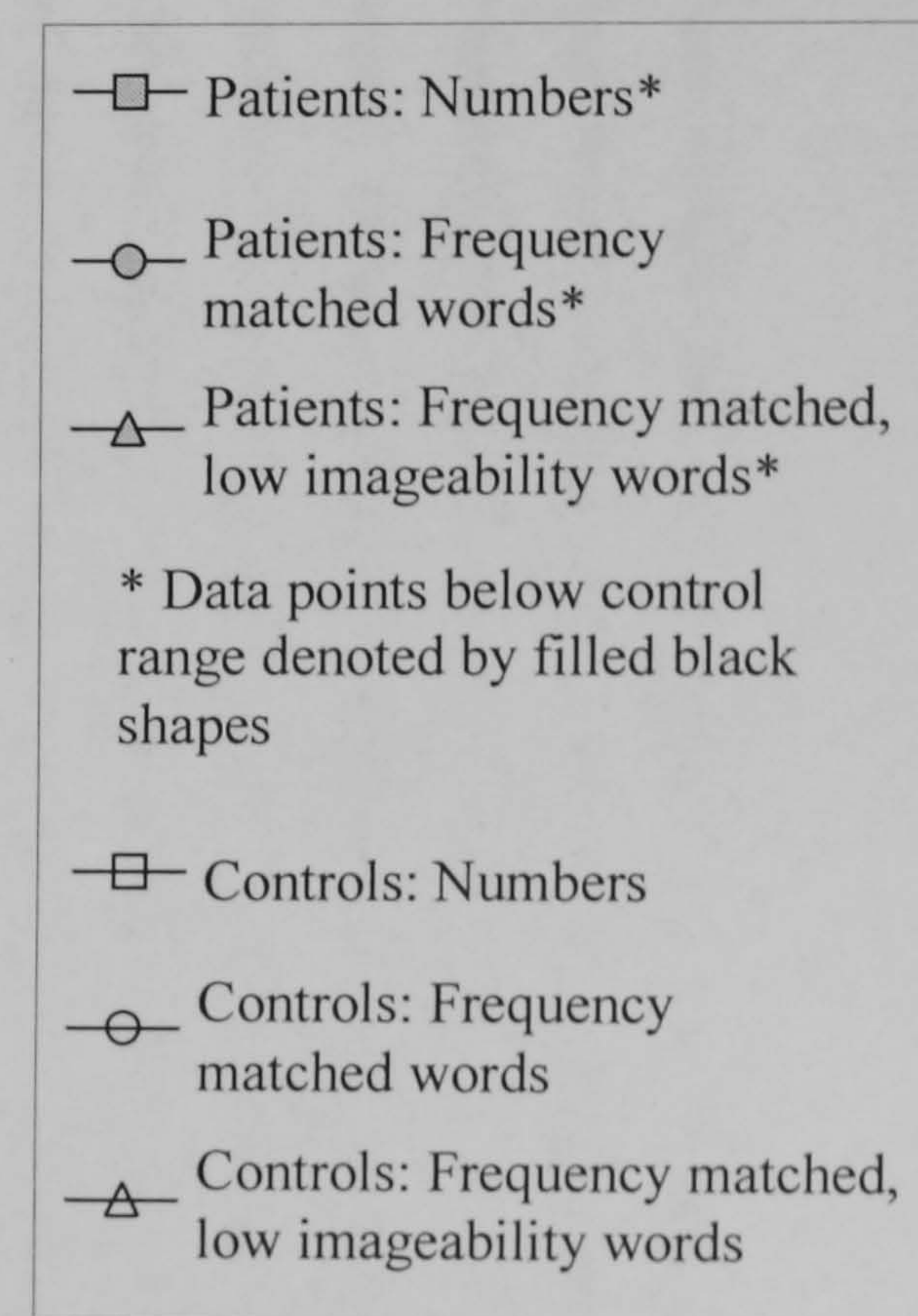
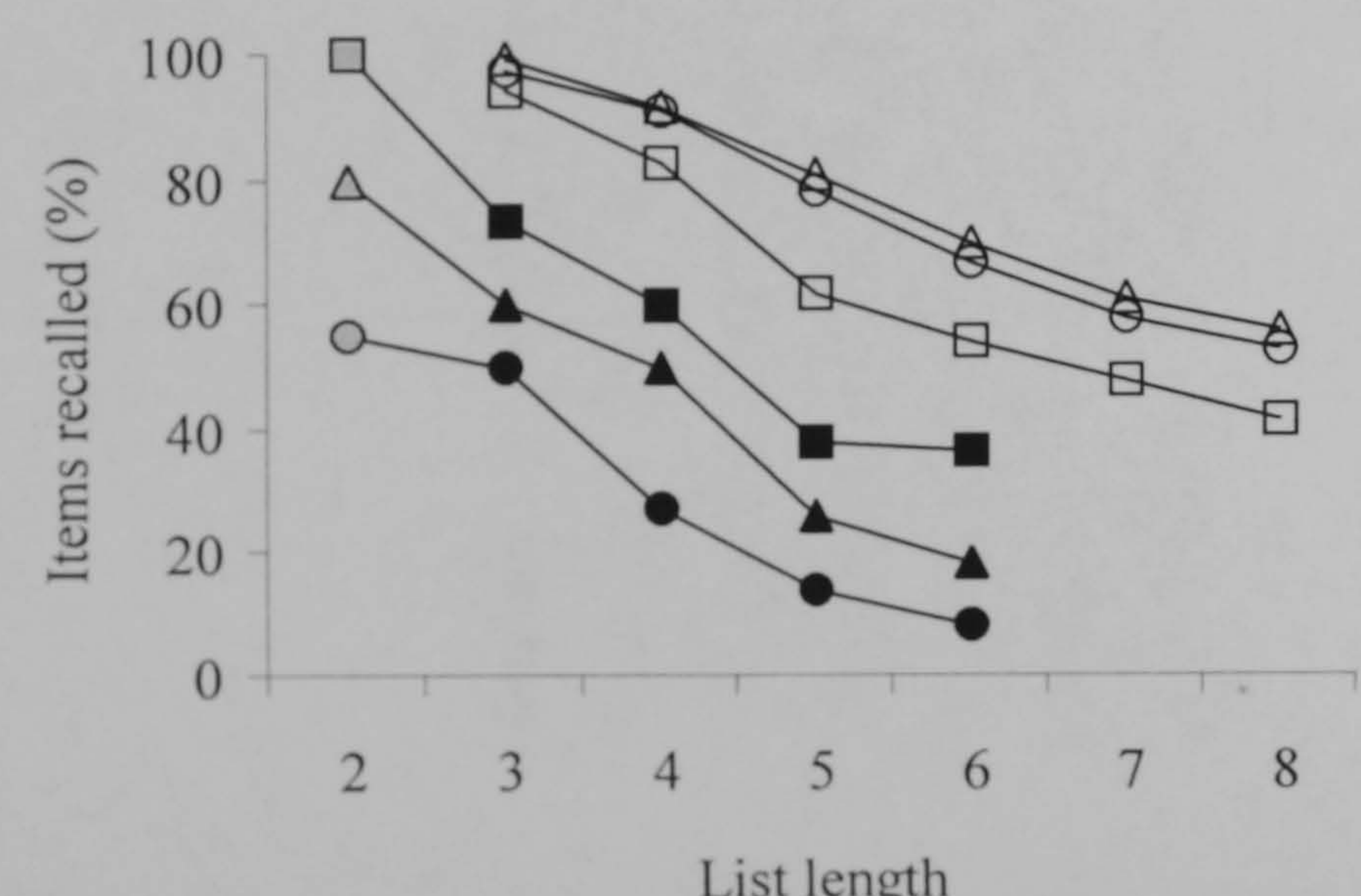


Fig. 3.2c. MK's recall of low frequency numbers and words



ISR for the numbers was within the normal range for GT, whereas EK was mildly impaired (her recall was lower than the control mean, and dipped below the lowest control score on a few list lengths). MK was more substantially below the normal range for number words. However, all three patients showed a much greater impairment in their recall of non-number words. EK and MK were very severely impaired on both sets of non-number words, with recall falling substantially below the normal range on every length tested. Recall of the non-number words was below the normal range for GT in parts of the data set. Moreover, all three patients showed an advantage for number over word recall that was much larger than the maximum observed in the controls (3%). The maximum advantage for number over word recall was 19% for EK, 30% for GT and 33% for MK.

MK recalled the high imageability words (set 2) more accurately than the low imageability words (set 1) ($t(97) = 2.63$, $p < 0.01$). Neither EK nor GT showed a significant difference between the two sets of words (EK: $t(118) = 1.38$, n.s.; GT: $t(117 = 1.57$, n.s.). However, in all three patients, the difference between the two sets of words was larger on some lengths than the maximum observed in control participants (14%), suggesting an enhanced effect of imageability in the ISR of these patients. This result is consistent with previous findings (Knott et al., 1997, see also Chapter 5).

3.3.2.2 Errors committed on numbers and words

Errors were analysed in the same way as for Experiment 1 but with one difference. For the low frequency multi-digit number lists, both the patients and controls sometimes produced number words that were not in the set, but shared 50% of their phonemes with a target number. These errors met the criterion for a phonological error because of the phonological overlap between numbers like 'thirteen' (in the set) and 'sixteen' (not in the set). However, they did not appear to result from the migration, substitution, addition or deletion of phonemes, and were therefore placed in a separate category of number intrusions from outside the set.

Table 3.3: *Errors on low frequency numbers and words for patients and controls in Experiment 2*

	EK			GT			MK			Control mean (max)		
Condition	Num	W1	W2	Num	W1	W2	Num	W1	W2	Numbers	W1	W2
Span	3	2*	3	3	2*	3	2*	1*	2*	3.67 (3-4)	4.1 (3-5)	4.3 (3-5)
No. items	120	90	120	120	90	120	90	60	90	140 (120-150)	152.5 (120-180)	160 (120-180)
Phonological	0	19*	8*	4*	17*	26*	4*	22*	24*	0.2 (2)	2.8 (12)	0.4 (2)
Unrelated	0	0	1	0	1	5*	1	2	3*	0.2 (1)	1.0 (3)	0.1 (1)
Omission	7	0	15	6	2	1	4	0	8	9.3 (17)	15.2 (29)	13.9 (30)
Order	6	1	9	1	0	1	0	0	0	8.8 (14)	8.8 (19)	10.0 (17)
Repetition	2	0	1	0	0	1	0	0	1	3.7 (8)	3.6 (10)	3.8 (10)
Within set intrusion	12	0	2	10	0	2	13	0	0	12.8 (19)	6.6 (17)	9.1 (21)
Number intrusion	12	–	–	5	–	–	5	–	–	10.3 (17)	–	–

* denotes abnormal performance. Figures refer to the total number of errors.

Num = numbers. W1 = frequency-matched words. W2 = frequency-matched, high imageability words.

Table 3.3 gives the number of errors made by the patients and controls in each category, combining across three list lengths: span, span +1 and span +2 items. The controls' recall of the number words was characterised by frequent intrusions, from both inside and outside the set, and by omission and order errors. They made a similar pattern of errors on the low frequency words, although within-set intrusions were less numerous than for numbers (word set 1: $t(11) = 5.67$, $p < 0.001$; word set 2: $t(11) = 2.61$, $p < 0.05$). In addition, there were more omission errors ($t(11) = 3.11$, $p < 0.05$), phonological errors ($t(11) = 2.22$, $p < 0.05$) and unrelated errors ($t(11) = 2.42$, $p < 0.05$) on the lower imageability words (set 1) compared with the numbers.

The patients' errors on the non-number words were different in nature to those made by controls. The number of phonological errors greatly exceeded the normal range on both sets of non-number words for all three patients. In addition, the patients with more severe semantic impairments made larger numbers of phonological errors. In contrast, the numbers of omission, order, repetition, and intrusion errors did not exceed the normal range. The number of unrelated errors also exceeded the normal range for GT and MK in the second set of non-number words. The patients' errors on the multi-digit number words were more similar to those made by controls. The numbers of omission, order, repetition and intrusion errors (from within and outside the set) did not exceed the normal range. However, the more severely impaired patients, GT and MK, made slightly more phonological errors on the number words than the controls. Evidence is presented below to suggest the patients' comprehension of low frequency multi-digit numbers was impaired, consistent with an association between semantics and phonological errors in ISR within the number domain.

As in Experiment 1, there were reliable differences between the pattern of errors in number and word recall for every patient. Out of six possible contrasts between the patients' pattern of errors in number vs. word recall (numbers vs. word set 1 and numbers vs. word set 2 for each of the three patients), all six revealed a statistically reliable difference at $p < 0.001$, with chi-squared values ranging from 27.3 to 57.1. The standardised residuals for phonological errors on the non-number words were high in every case.

3.3.3 Discussion

Two patients were mildly impaired at recalling the low frequency multi-digit numbers relative to controls, but all three patients were much more impaired at recalling the matched non-number words, making the difference between the materials greater in the patients than controls. Moreover, the quantity and quality of the patients' errors were similar to those of the control participants on the number words, but the patients made many more phonological errors than the controls on the non-number words. The patients' abnormal ISR advantage for numbers extended beyond single digits to low frequency multi-digit numbers. This is a potentially important finding because digit span can be preserved in patients with otherwise severe aphasia (Cohen, Verstichel, & Dehaene, 1997), suggesting that the ability to repeat digits may be over-learned or automated and therefore protected. It seems unlikely that low frequency numbers could be automated in the same way and therefore this possibility does not provide an adequate account of the data.

3.4 Experiment 3: Immediate serial recall of numbers and face-part words

A third experiment examined recall of middle frequency, mostly multi-digit numbers and frequency-matched words that loosely fitted into the category of 'face or head parts' (for example, mouth, fringe, beard). This study had several aims. First, comprehension of the number and face-part words could be directly compared in naming and word-picture matching tasks, making it possible to investigate whether the ISR difference between number and non-number words corresponded to a difference in comprehension. The results of these semantic tasks are discussed in a separate section below. Secondly, both the number and face-part words were drawn from closed semantic categories. In the experiments above, the number words were drawn from a single semantic category, whereas the non-number words were drawn from many semantic categories, making the non-number words less predictable. This experiment examined whether the superior recall of number words would persist after matching for this feature. Thirdly, it could be argued that the numbers in Experiment 1 were more imageable than the words they were matched with, because it is apparently easier to form a mental image of an Arabic numeral, e.g., '3', than a word with an intermediate imageability rating, e.g., 'small'. However, the highly imageable face part words used in this experiment were, according to published ratings, considerably *more* imageable than the digits 1 to 9. Consequently, if the advantage for number words persists in this experiment, it is unlikely to result from enhanced imageability effects in the patient group.

3.4.1 Method

Twelve number words (whole words rather than compounds) were compared with twelve face-part words in an ISR task. The items were matched as closely as possible for frequency using the Celex database (Baayen et al., 1993). It did not prove possible to match the groups for syllable length; the number words were significantly longer. The face-part words also had higher imageability ratings in the MRC online corpus (Coltheart, 1981). However, these two differences should reduce the recall advantage for number words shown by SD patients. The items are reproduced in Appendix 7.

EK was tested on lists containing three to seven items. GT and MK were tested on lists containing two to seven items. MK was additionally tested on a single face part word. The twelve controls were tested on lists containing three to eight items. Ten lists were tested at each length. In list construction, the number and face-part words were yoked so that matched items appeared in the same position in the lists. The numbers and face-part words were tested in blocks using an ABBA design.

3.4.2 Results

3.4.2.1 Recall accuracy

Figure 3.3 shows the percentage of number and face-part items recalled by the patients and controls, and Table 3.1 indicates span for these materials. The controls recalled the face-part words better than the numbers, consistent with the shorter length and higher imageability of the items in the former set ($t(11) = 7.94$, $p < 0.001$). In contrast, the patients showed better recall of the numbers than the face-parts (EK: $t(92) = 3.24$, $p < 0.01$; GT: $t(114) = 4.51$, $p < 0.0001$; MK: $t(116) = 7.40$, $p < 0.0001$). These analyses collapsed across list length. Recall of the number items was within the normal range for all three patients on every length tested. In contrast, recall of the face-part words was below the normal range for all three patients, on almost every length tested. Moreover, all three patients showed an advantage for number over word recall that was much larger than the maximum observed in the controls (32% for EK, 38% for GT and MK, 5% for the controls).

3.4.2.2 Errors on number and face-part words

Errors were categorised as for Experiment 1, with the additional category of 'outside-set intrusions'. For numbers, these were numbers not included in the set, and for face-part words, these were parts of the head or face not included in the set. Errors were placed in these categories even if they also met the criteria for a phonological error. Table 3.4 shows the number of errors of each type produced by patients and controls, combining across span, span +1 and span +2 list lengths.

The controls largely made intrusion, omission and order errors for both numbers and face-parts. Several error types were more numerous for face-parts than for numbers, including omissions ($t(11) = 2.38$, $p < 0.05$), order errors ($t(11) = 3.53$, $p < 0.01$), repetitions ($t(11) = 3.02$, $p < 0.05$) and within-set intrusions ($t(11) = 2.94$, $p < 0.05$). Intrusions from outside the set were more numerous for the numbers ($t(11) = 10.19$, $p < 0.001$).

Figures 3.3a-c: *Recall of face parts and matched number words by patients and controls in Experiment 3*

Fig. 3.3a. EK's recall of numbers and face parts

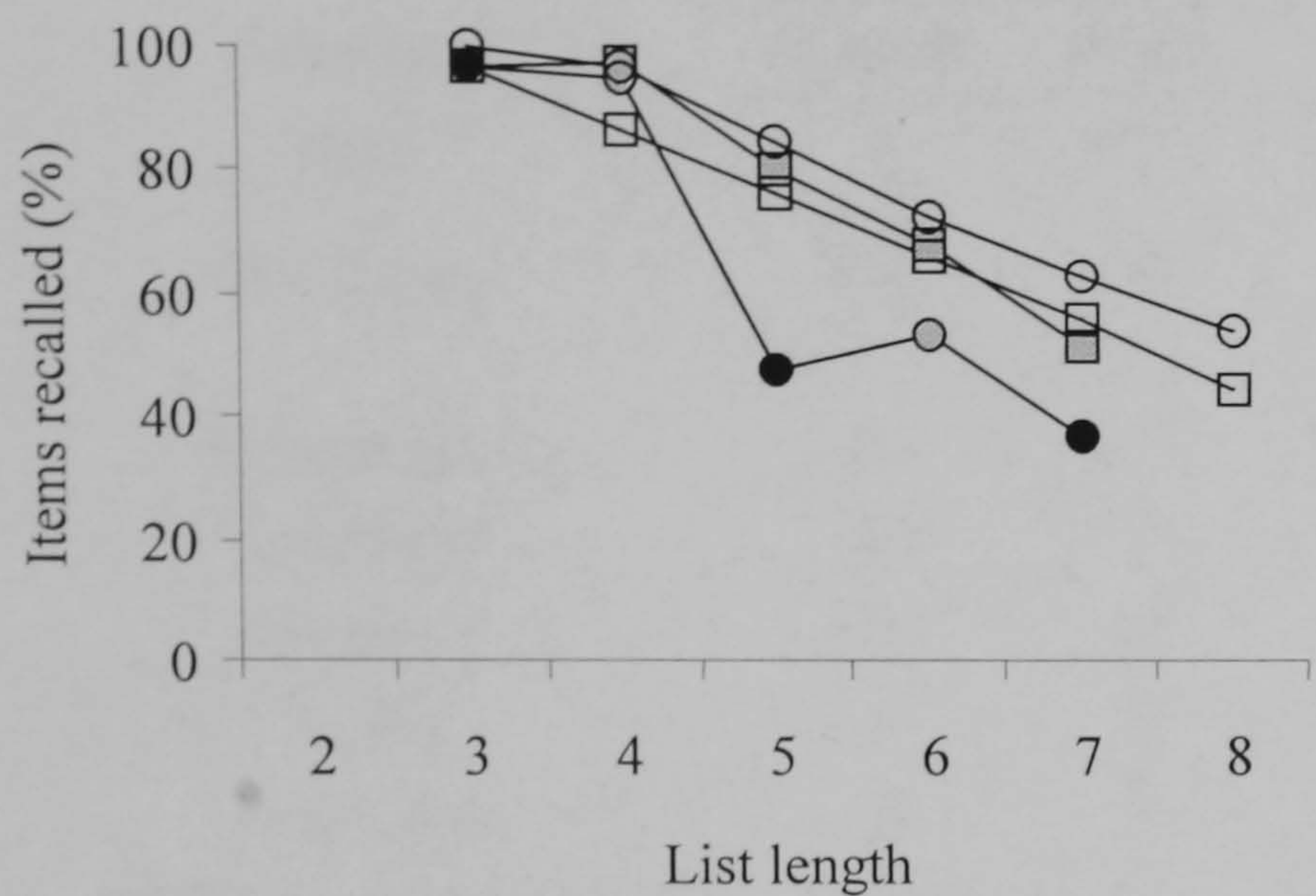
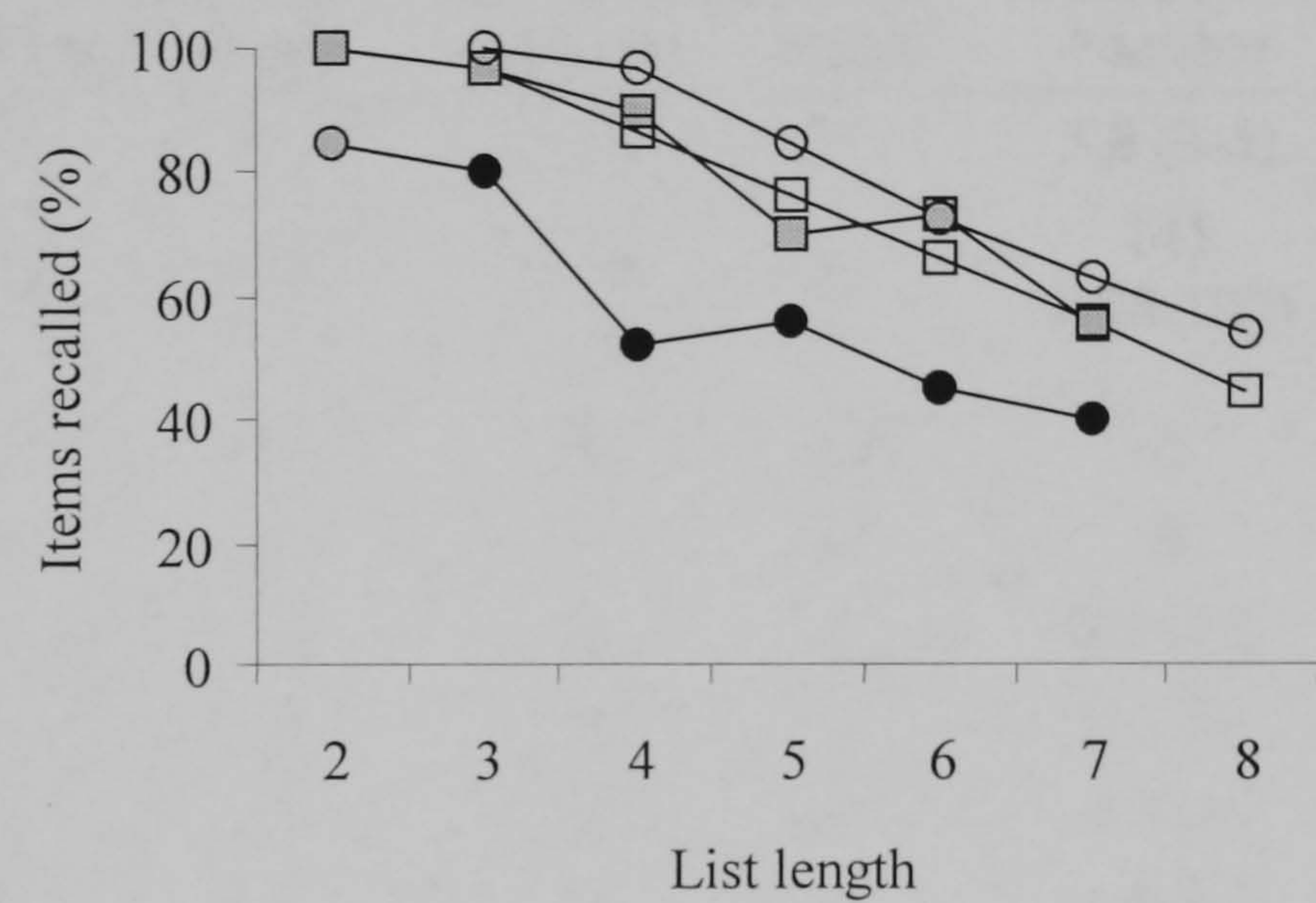


Fig. 3.3b. GT's recall of numbers and face parts



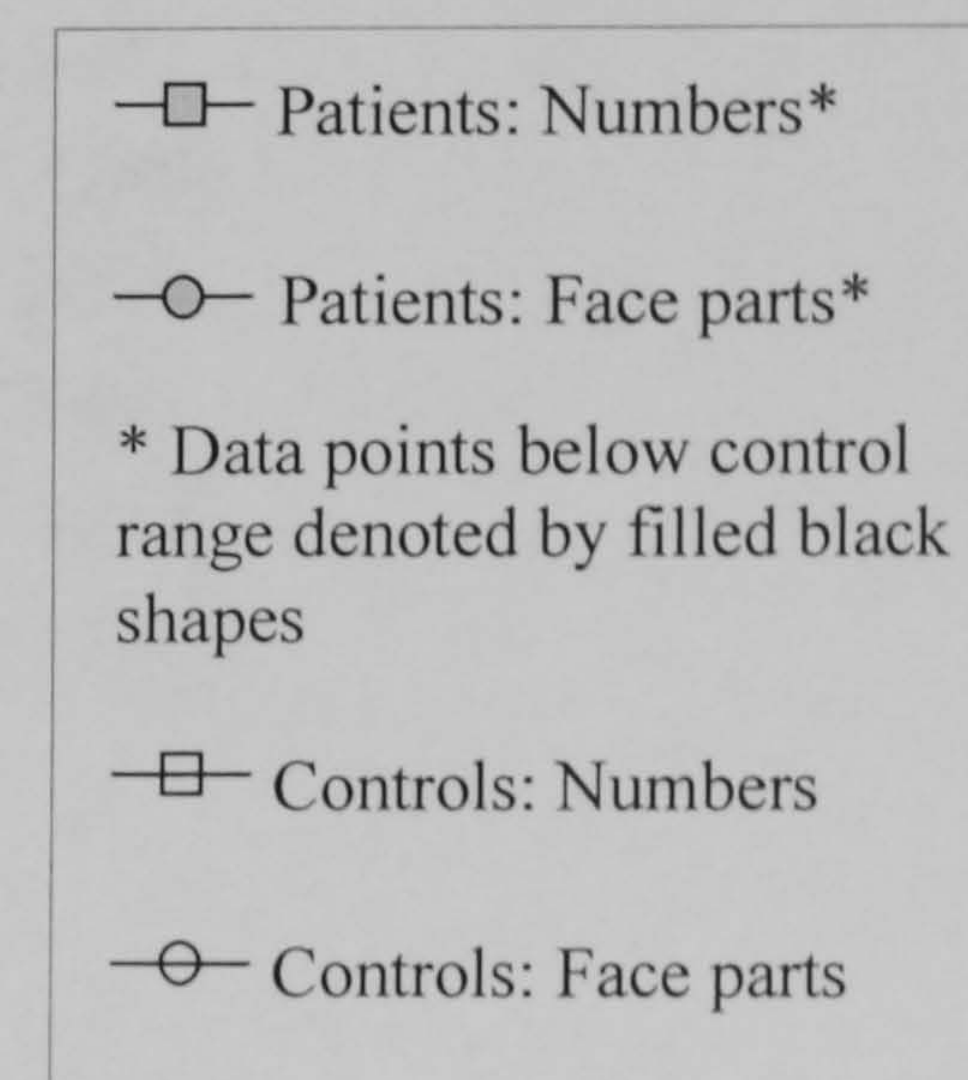


Fig. 3.3c. MK's recall of numbers and face parts

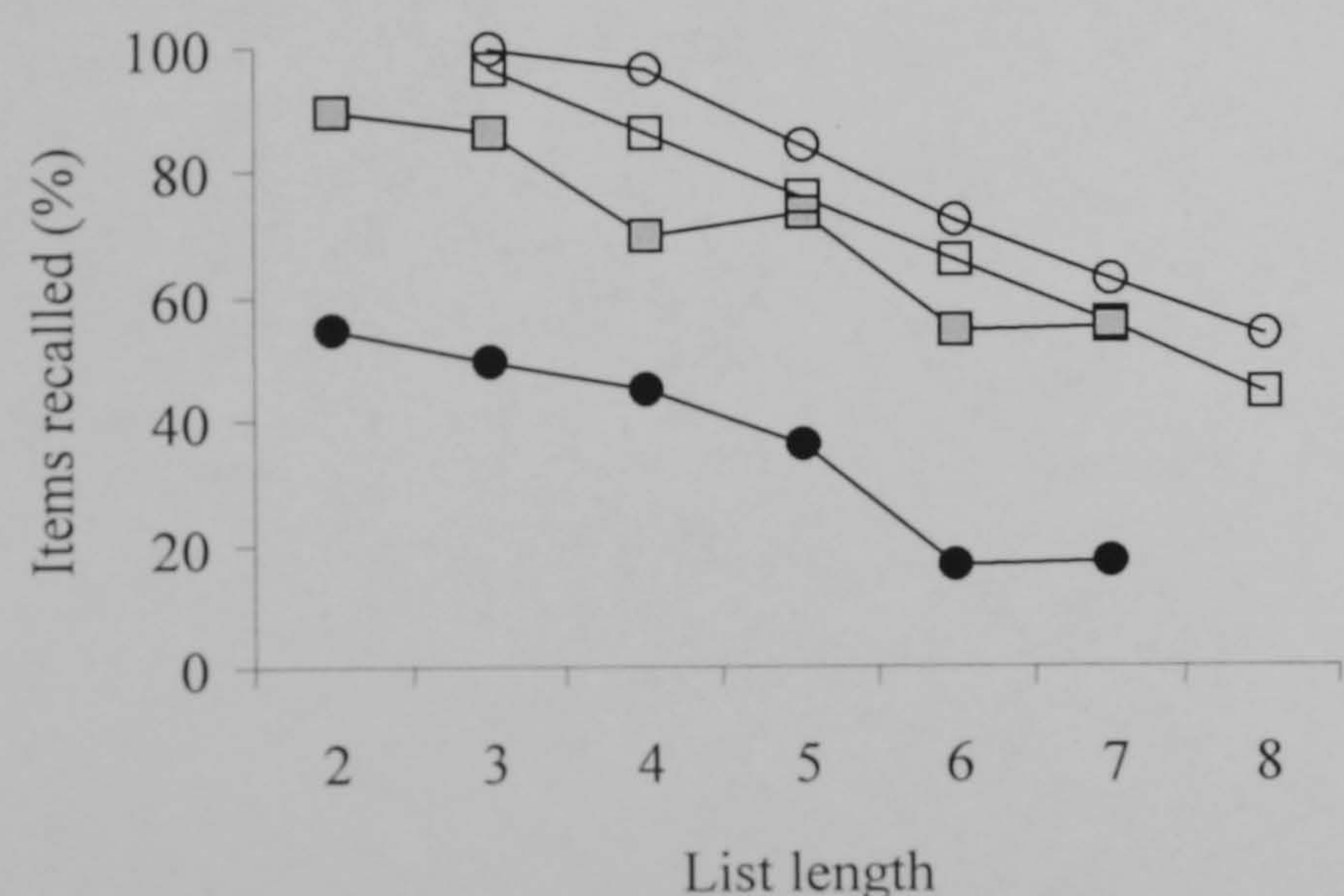


Table 3.4: *Errors on numbers and face-part words for patients and controls in Experiment 3*

	EK		GT		MK		Control mean (max)	
Condition	Number	Word	Number	Word	Number	Word	Number	Word
Span	4	3*	4	3*	3	1*	3.8 (3-5)	4.75 (4-6)
No. items	150	120	150	120	120	60	145 (120-180)	172.5 (150-210)
Phonological	1	17*	1	34*	0	15*	0	0.1 (1)
Unrelated	0	1	0	12*	0	9*	0	0.1 (1)
Omission	9	15	4	0	4	0	8.3 (19)	15.8 (44)
Order	3	5	5	0	4	0	4.1 (9)	9.8 (18)
Repetition	2	0	3	0	1	0	1.6 (5)	3.6 (9)
Within set intrusion	13	1	14	0	7	0	9.3 (21)	14.1 (25)
Outside set intrusion	10	0	8	0	14	0	10.5 (17)	0.1 (1)

* denotes abnormal performance. Figures refer to the total number of errors.

The three patients had a virtually normal pattern of errors for the number words. Omission, order, repetition, within and outside-set intrusion and unrelated errors were within the normal range. The patients' errors on the face-part words were different in nature. All three patients made a large number of phonological errors that greatly exceeded the normal range, and the number of unrelated errors was also above the normal range for GT and MK. In contrast, the numbers of omission, order, repetition and intrusion errors were within the normal range.

Error patterns on the number and face-part words were significantly different for all three patients (EK: $\chi^2(1) = 42.51, p < 0.001$; GT: $\chi^2(1) = 85.79, p < 0.001$; MK: $\chi^2(1) = 58.85, p < 0.001$). The standardised residuals were high for phonological errors on the face-part words in every case, suggesting that the large number of phonological errors on these words underpinned the difference in error patterns.

3.4.3 Discussion

As in the previous two experiments, the patients recalled the number words at relatively normal levels, but were significantly impaired at recalling the non-number face-part words. In addition, the patients made many more phonological errors than the controls on the non-number words but not on the number words. The difference between number and non-number words is reminiscent of the difference between known and degraded words observed in Chapter 2 and previous studies (e.g., Patterson et al., 1994). The ISR difference between number and non-number words remained in this experiment when both sets of items were drawn from closed semantic categories. However, the category of 'face parts' may have been less salient for the patients if their comprehension of the face-part words was impaired. The control participants may have been better able to use the face-part category to guide recall, consistent with their larger number of within-set intrusion errors.

3.5 Experiment 4: Immediate serial recall of letters

Experiments 1 to 3 suggest that number repetition is relatively intact in SD compared with word repetition, even when the materials are broadly equated for word length, frequency, imageability, set size, and open or closed semantic categories. However, there are clearly other differences between number and non-number words that could contribute to the better recall of number than non-number words: for example, numbers can be represented by single characters and occur in a sequence. Moreover, in some neurological patients, knowledge of number sequence appears to mirror comprehension of other ordered series, such as days of the week or months of the year (Cipolotti et al., 1991; Dehaene & Cohen, 1997; Thioux et al., 1998). Letters appear to share some of these unusual properties of numbers. Therefore, a fourth experiment examined ISR for letters, and compared letters with the single-digit numbers and high frequency non-number words used in Experiment 1. This experiment also compared phonologically similar and dissimilar letters. If SD patients have intact phonological STM capacities, as is generally assumed, they should show a normal effect of phonological similarity in ISR (Knott et al., 2000, see also Chapter 4). Normal participants show poorer recall of phonologically similar than dissimilar items (e.g., Conrad & Hull, 1964); an effect which is typically attributed to phonological coding in STM.

3.5.1 Method

Following Knott et al. (2000), the patients were asked to repeat letters from the phonologically similar set E, C, T, P, V, B, G, D or the phonologically dissimilar set S, Q, Y, R, J, F, W, L. The set size was limited to eight items because there are only eight phonologically similar letters. This is similar to the set size of nine used in Experiment 1. The patients and twelve control participants were tested on lists of four and six letters. There were twenty lists at each length divided equally between the phonologically similar and dissimilar sets. The similar and dissimilar letters were presented in blocks using an ABBA design. They were read aloud by the experimenter at a rate of one item per second.

3.5.2 Results

Table 3.5 shows the number of phonologically similar and dissimilar letters recalled in lists of four and six items. For comparison purposes, Table 3.5 also gives the number of single-digit number words and high frequency non-number words (second set) recalled in Experiment 1.

Table 3.5: *ISR for letters, digits and frequency-matched words*

	EK	GT	MK	Controls Mean (range)
4 similar letters	90	85	73	82.7 (63 – 98)
4 dissimilar letters	100	90	83*	96.5 (85 – 100)
4 digits	100	100	100	99.6 (95 – 100)
4 words (set 2, Exp. 1)	95	83*	68*	97.5 (87 – 100)
6 similar letters	68	75	53	70.1 (53 – 82)
6 dissimilar letters	87	82	72	83.8 (72 – 97)
6 dissimilar – similar letters	18	7	18	13.6 (5 – 23)
6 digits	90	100	90	88.7 (70 – 100)
6 words (set 2, Exp. 1)	60	53*	48*	74.0 (55 – 90)

Note: Figures indicate percentage of items recalled in correct order

* denotes abnormal performance

3.5.2.1 Comparison of letter, digit and word recall

For the normal participants, letter recall was intermediate between ISR for single-digit numbers and non-number words. The controls recalled the single-digit numbers better than both the phonologically similar letters ($t(11) = 8.37$, $p < 0.0001$) and the phonologically dissimilar letters ($t(11) = 2.81$, $p < 0.05$). Their recall of high frequency words (Experiment 1, second set) was poorer than ISR for phonologically dissimilar letters ($t(11) = 3.53$, $p < 0.01$) but better than their performance with phonologically similar letters ($t(11) = 4.15$, $p < 0.01$). The patients showed a comparable pattern. All three patients recalled the single-digit numbers at a higher level than the phonologically similar letters (EK: $t(32) = 3.71, p < 0.001$; GT: $t(19) = 5.90, p < 0.0001$; MK: $t(35) = 7.84, p < 0.001$). Two of the patients also recalled the single-digits at a higher level than the phonologically dissimilar letters (GT: $t(19) = 3.90, p < 0.001$; MK: $t(33) = 4.02, p < 0.001$; EK: $t(38) < 1$). In addition, in every case, the phonologically dissimilar letters were recalled more accurately than the non-number words (EK: $t(27) = 2.80, p < 0.01$; GT: $t(35) = 2.96, p < 0.01$; MK: $t(36) = 3.31, p < 0.01$). GT showed better recall of phonologically similar letters than non-number words ($t(34) = 2.04$, $p < 0.05$), whereas EK and MK showed no difference between these conditions (EK: $t(32) < 1$; MK: $t(34) = 1.01$, n.s.). These analyses combined data from both list lengths.

Table 3.5 indicates that EK and GT showed normal recall of phonologically similar and dissimilar letters. Therefore, these two patients had intact digit and letter repetition abilities, but impaired word repetition abilities. The most severely impaired patient, MK, had mildly impaired letter repetition abilities in the context of severely impaired word repetition and normal digit repetition. MK, unlike the other two patients, was impaired on the degraded letters sub-test in the VOSP. She was only able to name ten out of twenty degraded letters, suggesting that she may have had difficulty recognising their visual forms. MK was also unique in that she made a number of phonological errors and non-letter intrusions in the letter span task (for example, she recalled G as 'chee', and she recalled Q as 'car' and then changed it to 'R'). Therefore, it seems that letter repetition may have been intact in EK and GT because their knowledge of letters was relatively intact, whereas MK's letter repetition may have been impaired because her knowledge of letters was degraded.

3.5.2.2 Phonological similarity effects

The controls showed a highly significant effect of phonological similarity ($t(11) = 7.94, p < 0.0001$), combining across list lengths. EK and MK also showed better recall of phonologically dissimilar items ($t(35) = 3.33, p < 0.01$ and $t(37) = 2.86, p < 0.01$ respectively). However, the numerical difference between phonologically similar and dissimilar items failed to reach significance for GT ($t(38) = 1.17$, n.s.). The recall difference between phonologically similar and dissimilar letters was within the normal range for all three patients (see Table 3.5).

3.5.3 Discussion

EK and GT showed relatively normal recall of letters, but MK, a more severe patient, was impaired. MK also exhibited poor comprehension for letters, possibly accounting for her impaired recall. Therefore, MK showed intact digit span in the context of impaired letter span and severely impaired word span, whereas EK and GT showed intact recall of single-digit numbers and letters but impaired word span. These data are consistent with the notion that ISR is relatively preserved for items that form an ordered series. It is possible that ordered sequences such as letters and numbers are relatively preserved in SD because they derive support from a intact understanding of spatial relations (see Gevers, Reynvoet, & Fias, in press).

All three patients exhibited better recall of phonologically dissimilar than similar letters, although the size of the phonological similarity effect may have been slightly reduced in GT. The presence of a phonological similarity effect is consistent with the suggestion that SD patients use normal phonological encoding in STM, and that semantic rather than phonological impairments are the cause of their poor recall. This argument applies equally to the non-number words examined here and the semantically degraded words studied in Chapter 2 (see Chapter 4 for further discussion). This finding also suggests that any differences in phonological similarity between the number and non-number words used in Experiments 1 to 3 should have had comparable effects on the ISR of the patients and controls.

3.6 Understanding of number

Experiments 1-4 suggest that SD patients have relatively intact ISR for numbers (and possibly other ordered series) and more impaired ISR for non-number words. Substantial and reliable ISR differences were observed for every patient, and these were largest in the patient with the most severe semantic impairment. The ISR difference was not eliminated by matching for frequency, imageability, word length, set size or open vs. closed semantic category, suggesting that these factors cannot account for the pattern of results. There are likely to be other differences between numbers and words that cannot be eliminated through the matching of materials. These differences, however, should affect controls as well as patients, and do not readily explain the pattern of impaired word recall but intact number recall that is observed in SD. This section examines the patients' knowledge of numbers to investigate the possibility that ISR for these items is specifically preserved because comprehension of numbers is relatively intact in SD.

There were four elements to these investigations of number processing. First, naming and word-picture matching tasks were devised for the number and face-part words used in Experiment 3, to establish if the superior ISR for numbers corresponded to better comprehension. Secondly, the patients' ability to transcode between Arabic numbers and spoken number words was assessed for the items used in Experiments 1 to 3. Thirdly, the patients' understanding of the numbers used in the first three experiments was assessed using sequence and magnitude judgement tasks, providing a means of evaluating the idea that their poorer recall of lower frequency multi-digit numbers corresponded with poorer comprehension. Finally, the patients' abilities to understand and manipulate numbers were explored more generally. On most of these tests, the performance of controls was at ceiling, and therefore only one age-and education-matched control was tested for each patient unless otherwise stated.

3.6.1 Comprehension of number and face-part words

3.6.1.1 Method

Naming and word-picture matching tests were devised for the number and face-part words used in Experiment 3. The numbers were represented pictorially as dots; tens were depicted as clusters of ten dots, and units were depicted as single dots. For example, the word 'thirteen' was shown as a cluster of ten dots and three individual dots. The patients were told that the clusters contained ten dots. Each set of dots was depicted within the same sized box. In naming, the patients were asked to provide a name for each dot picture in turn. In word-picture matching, the patients were shown all twelve pictures together and were asked to point to the one that represented a particular number word. The face-part tests used a complete picture of a face. In naming, the patients were asked to name the face-parts that were indicated by arrows. In word-picture matching, the patients were asked to select the arrow that pointed to a particular face-part. The item 'brow' was omitted from these tests, as it was not pictorially distinct from 'forehead'. This item was replaced by another face-part in the word-picture matching test, so that the number of distracters remained the same for numbers and face-parts.

3.6.1.2 Results

The three controls performed without error on both the number and face-part words. Table 3.6 gives the results for the patients. All three patients were virtually at ceiling on naming and word-picture matching with number words. EK and GT performed these tasks without any difficulties. MK made one error on the word-picture matching numbers test, which apparently resulted from confusion of the words 'seventeen' and 'seventy'. In contrast, all the patients were impaired on the face-parts. In naming, their errors were predominantly 'don't know' responses (EK: 4/5 errors, GT: 3/5 errors, MK: 3/8 errors). GT and MK also made semantic errors (e.g., forehead → 'chin'; GT: 2/5 errors, MK: 1/8 errors). MK made two responses that indicated partial phonological knowledge about the target (neck → 'something n...' and nose → 'nuv?'). The patients performed significantly better with the numbers than with the face-parts when the results for naming and word-picture matching were combined (EK, $\chi^2(1) = 6.71$, $p < 0.01$; GT, $\chi^2(1) = 5.32$, $p < 0.05$; MK, $\chi^2(1) = 15.87$, $p < 0.0001$).

Table 3.6: *Naming and word-picture matching using the number and face-part words from Experiment 3*

	Max	EK	GT	MK
Naming: numbers	12	12	12	12
Naming: face-parts	11	6*	6*	3*
Word-picture matching: numbers	12	12	12	11*
Word-picture matching: face-parts	11	9*	10*	5*

* denotes abnormal performance

3.6.1.3 Discussion

There was a comprehension difference between the number and face-part words, even though these items were matched for frequency, suggesting that the number domain may be relatively intact in SD. The ISR difference between number and face-part words corresponded to a difference in comprehension, and patient MK who had the poorest ISR for face-parts also achieved the lowest scores for naming and comprehending items from this category. This pattern of results is consistent with the suggestion that the ISR difference between number and non-number words is an instance of the ISR difference between known and degraded words.

3.6.2 Transcoding of single-digit and lower frequency multi-digit numbers

3.6.2.1 Method

The patients were asked to read aloud and write Arabic numerals for each of the number words used in Experiments 1 to 3. An inability to perform these transcoding tasks might indicate that the patients did not comprehend the words used in the ISR experiments. There is controversy, however, about the extent to which transcoding tasks rely on 'semantic' representations of number. In some models of numerical processing, an internal abstract representation of numerical magnitude plays a critical role in transcoding tasks (McCloskey, Caramazza, & Basili, 1985), whereas other models have proposed that there are additional non-semantic transcoding routes (Cipolotti & Butterworth, 1995; Dehaene, 1992; Dehaene & Cohen, 1995; Deloche & Seron, 1982a).

3.6.2.2 Results

The results are summarised in Table 3.7. The controls performed almost perfectly. The patients were able to translate between spoken number words and Arabic numerals almost without error; only the items 'billion' and 'trillion' created difficulties. The milder patients showed some understanding that these words represented large numbers (EK wrote trillion as 100 and billion as 10000; GT wrote billion as 500 but failed to produce a response for trillion; MK failed to respond on either item). The control participant who did not write billion and trillion accurately made errors that were much closer to the correct response.

3.6.3 Comprehension of single-digit and lower frequency multi-digit numbers

3.6.3.1 Method

Two tests examined comprehension of the single-digit number words used in Experiment 1 relative to the lower frequency multi-digit numbers used in Experiments 2 and 3. First, the patients were asked to arrange cards with the number words printed on them in numerical order. The number words were read aloud by the experimenter throughout the test. Secondly, the patients were asked to select the number out of four that was numerically closest to a target. For example, they were asked 'Which number is nearest to five: eight, nine, two or three?'. The correct response could be either larger or smaller than the target. The alternatives were drawn from the same experimental set of numbers as the target. The numbers were simultaneously read aloud and presented as written number words during the test. This 'which number is nearest' test has some similarities with number comparison tasks that require participants to make judgements about which of two numbers is larger. Healthy participants' reaction times in comparison tasks decrease as the numerical distance between the numbers to be compared is increased; i.e., participants are faster to say which is bigger between '2 and 9' than between '2 and 3'. In addition, for equal distances, comparison times are slower for larger numbers (Dehaene, 1989; Dehaene, Dupoux, & Mehler, 1990; Moyer & Landauer, 1967). Comparison times are thought to fit a compressive function approximating Weber's law (Dehaene, 1992). In line with this law, the numerical distances in this test between the target and the choices, expressed as a proportion of the size of the target, did not differ significantly between the single-digit numbers used in Experiment 1 and the predominantly multi-digit numbers used in Experiment 3. The average distance between the target and choices was, however, larger for the low frequency multi-digit numbers used in Experiment 2, largely because of the distorting influence of the items 'billion' and 'trillion' (although if anything, this should have made these items easier).

3.6.3.2 Results

The results are summarised in Table 3.7. The patients were able place the digit words in the correct order, but unlike controls, they made some errors on the mid-frequency numbers from Experiment 3, and a larger number of errors on the low frequency numbers from Experiment 2. Some of the patients' errors appeared to result from confusions between '*-teen*' words like 'thirteen' and 'fourteen', and '*-ty*' words like 'thirty' and 'forty' (e.g., 16 → 70 → 18) but some did not (e.g., 80 → 90 → 70 → 19). MK was very slow at this task and appeared to be using an ineffective 'counting up' strategy that was successful with single-digit numbers but not larger numbers.

The patients performed even more poorly on the 'which number is nearest' task, perhaps because it involved calculation as well as an understanding of numerical magnitude (see below). EK and MK made errors on this task even when it involved single-digit numbers. It seems unlikely that the patients' errors were due to a failure to understand the instructions, because they largely selected the distracter second nearest to the target for single-digit numbers and medium frequency numbers (9/12 errors across the patients). This error pattern suggests that they knew, at least approximately, about the sequence of numbers and their magnitudes. In contrast, they selected the approximate distracter less often for low frequency multi-digit numbers (2/11 errors across the patients).

Table 3.7: *Semantic tests for the three sets of number words used in Experiments 1 to 3*

		Exp. 1: Single-digit	Exp. 2: Low frequency multi-digit	Exp. 3: Medium frequency, mostly multi-digit
	Max	9	9	12
EK	Reading numerals	9	7	12
	Writing numerals	9	7	12
	Ordering numbers	9	6*	11*
	Which number is closest?	8*	7*	8*
GT	Reading numerals	9	7	12
	Writing numerals	9	7	12
	Ordering numbers	9	8*	12
	Which number is closest?	9	7*	11*
MK	Reading numerals	9	7	12
	Writing numerals	9	7	12
	Ordering numbers	9	3*	10*
	Which number is closest?	7*	2*	8*
Controls	Reading numerals	9	9	12
	Writing numerals	9	7 – 9	12
	Ordering numbers	9	9	12
	Which number is closest?	9	9	12

* denotes abnormal performance

3.6.3.3 Discussion

There is evidence for an association between semantic knowledge and ISR within the numbers domain as well as between the number and non-number categories. The patients' understanding of lower frequency numbers was poorer than their understanding of single-digit numbers. Similarly, they showed poorer recall of these items and made more frequent phonological errors on them. The patients performed at ceiling on one task involving single-digit numbers but they made a larger number of errors in a second task that involved calculation. This finding is consistent with the view that an understanding of numerical magnitude is largely intact in SD but that knowledge of calculation procedures is impaired (Cappelletti et al., 2001). Some additional assessments of calculation ability that provide some support for this view are presented below.

3.6.4 Other tests of number comprehension

Some additional data about the patients' numerical abilities that did not relate specifically to the sets of numbers used in Experiments 1 to 3 were collected. First, the patients' number knowledge was examined using a numerical comparison task. In addition, their calculation abilities were assessed in two tasks; they were given arithmetic questions to solve (e.g., '11 + 8', '4 × 1') and they were asked to provide the next number in sequences like '4, 7, 10, 13, ?', where the number series itself specified the operation required to generate the next number (add 3). Manabu Ikeda kindly provided the data from the first two of these tasks. Although the primary focus of the current chapter is on the relationship between numerical understanding and phonological errors in ISR, these data are useful for evaluating the claim that numerical abilities are spared in SD.

3.6.4.1 Number comparison task

The number comparison task required patients to judge which of two numbers was numerically larger. There were twenty questions. Six involved comparisons between single-digit numbers (e.g., 9 and 2), two involved one vs. two-digit numbers (e.g., 7 and 13) and eight involved comparisons between two-digit numbers (e.g., 10 and 16). In addition, two involved two vs. three-digit numbers (e.g., 105 and 89) and two involved comparisons between three-digit numbers (e.g., 948 and 199). Four patients, including PD, were tested on this task. In every case, their performance was errorless suggesting that the patients' difficulties with the 'which number is nearest task' may have resulted from the fact that multiple comparisons and/or calculation were required. In addition, it is possible that the use of Arabic numbers rather than printed number words facilitated the patients' performance in this task.

3.6.4.2 Arithmetic questions

The four patients were asked to solve 108 calculations, written out on paper. They were allowed to write down their workings. There were 27 calculations for each of the mathematical operations (i.e., addition, subtraction, multiplication and division), presented in a mixed fashion. The particular operation required on each trial was indicated by the standard symbols, '+', '−', 'x', and '÷'. The meaning of these symbols was explained to each patient prior to testing and it proved necessary to provide repeated reminders throughout the test. The sums involved one-, two- and three-digit operands (see Table 3.8).

Incorrect responses were assigned to one of three categories: 1) 'no response' errors, 2) 'symbol comprehension' errors, which were responses that would have been correct if a different operation had been required (e.g., $21 - 9 = 30$: addition used instead of subtraction; $14 \times 21 = 35$: addition used instead of multiplication), and 3) 'other calculation errors', which did not fall within the first two categories. Some examples of errors from this heterogeneous group are provided below. Table 3.8 shows the number of responses in each of these categories for each patient as a function of mathematical operation and operand size. No control data is available for this test.

Table 3.8: *Calculation accuracy and error types*

			EK				GT				PD				MK			
	Type of sum	No items*	Cor	NR	Symb	Other	Cor	NR	Symb	Other	Cor	NR	Symb	Other	Cor	NR	Symb	Other
Add	1+1	8(7)	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
	11+1	10(9)	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
	11+11	9	100	0	0	0	100	0	0	0	89	0	0	11	100	0	0	0
	Overall	27(25)	100	0	0	0	100	0	0	0	96	0	0	4	100	0	0	0
Subtract	1-1	4	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
	11-1	10	90	0	0	10	100	0	0	0	90	10	0	0	90	0	10	0
	11-11	10(9)	90	0	0	10	100	0	0	0	78	0	0	22	60	10	0	30
	111-111	3(2)	100	0	0	0	100	0	0	0	50	0	0	50	100	0	0	0
	Overall	27(25)	93	0	0	7	100	0	0	0	84	4	0	12	81	4	4	11
Multiply	1x1	9(8)	67	0	11	22	100	0	0	0	63	0	13	25	44	0	22	33
	11x1	9	78	0	0	22	100	0	0	0	33	33	22	11	33	44	0	22
	11x11	9(8)	0	89	0	11	100	0	0	0	13	25	38	25	0	67	33	0
	Overall	27(25)	48	30	4	19	100	0	0	0	36	20	24	20	26	37	19	19
Divide	1/1	5	80	0	20	0	100	0	0	0	100	0	0	0	60	20	0	20
	11/1	4	75	0	0	25	100	0	0	0	25	75	0	0	25	25	50	0
	111/1	9(8)	11	56	0	33	100	0	0	0	13	50	0	38	0	67	22	11
	111/11	9(8)	33	44	0	22	100	0	0	0	25	38	0	38	22	67	0	11
	Overall	27(25)	41	33	4	22	100	0	0	0	36	40	0	24	22	52	15	11
Grand Mean		108(100)	70	16	2	12	100	0	0	0	63	16	6	15	57	23	9	10

Figures refer to percentage of item presented. Cor = correct, NR = no response, Symb = symbol comprehension error.

* The numbers of items shown in parentheses refer PD, who was tested on fewer items due to time constraints.

EK, GT and MK performed perfectly on the addition sums, whereas PD made a single error. The patients were also relatively good at subtraction. The less impaired patients made virtually no errors and the more severely impaired patients made a handful of errors on the multi-digit problems. EK, MK and PD showed much poorer performance on multiplication and division questions, scoring an average of only 37% and 33% respectively. In contrast, GT was able to carry out multiplication and division without difficulty, perhaps because of his premorbid vocational experience with number and calculation. The patients' performance was strongly affected by the size of the operands. As a group, they achieved 68% and 85% accuracy in the easiest multiplication and division questions involving two single-digit operands. In contrast, they only obtained 28% in questions that required two two-digit operands to be multiplied, and 45% in questions that required a three-digit number to be divided by a two-digit number.

The most frequent error was a failure to respond. The patients also repeatedly made errors that appeared to result from the selection of an inappropriate mathematical operation, suggesting that the patients may not have understood the meanings of the mathematic symbols. EK, MK and PD indicated that they understood the symbol '+', but did not comprehend '−', '×', or '÷'. GT comprehended all four symbols. The patients also made a considerable number of other calculation errors. Some of these errors were reminiscent of an impairment of arithmetic facts/rules, e.g., 2 x 3 = 8, 8 x 1 = 16 (McCloskey, 1992) and some may have resulted from failures to follow multi-digit procedures; e.g., 34 - 18 = 26, rather than 16. The patients were able to read the Arabic numbers from 1 to 20 without error, suggesting that these errors did not result from an inability to recognise Arabic numbers.

3.6.4.3 'Which number comes next' test

As the patients' failure to understand mathematical symbols appeared to contribute to their poor performance on the previous test, a calculation task was devised that avoided the use of such symbols. EK, GT, MK and twelve controls were tested on this 'which number comes next' task, which required the next number in a sequence to be calculated (for example, 4, 7, 10, 13, ?). The sequences themselves specified the operation that produced the answer, avoiding the need to use mathematical symbols. The numbers were presented in Arabic numerals and the patients wrote down their answers and their workings. There were forty sequences. In ten of them, a number was added to produce the answer, and in another ten, a number was taken away to produce the answer. In the next ten, the previous number was multiplied by a constant to produce the next number, and in the final ten, the previous number was divided by a constant to produce the answer. These different types of sequences were blocked and the patients were told that the blocks required different operations. Four of the twenty addition and subtraction sequences were 'second order' as the amount that was added or subtracted was changed by a constant amount each time.

The patients' performance on these sequences, shown in Table 3.9, was generally very good. GT was functioning at a particularly high level, perhaps because of his premorbid vocational experience with number and calculation. EK showed some impairment on sequences involving division, and MK showed some impairment on sequences involving subtraction, multiplication and division. These impairments did not seem to be accountable by a failure to understand the task or the numbers involved – MK showed perfect performance on the higher order sequences, demonstrating that she was able to detect underlying patterns and make inferences about the next number. Instead, the impairments could be explained by a poor understanding of the multiplication and division procedures.

Table 3.9: *Performance of patients and controls on the number sequence test*

	Max	EK	GT	MK	Control mean (range)
Addition	10	7	10	9	9.3 (7 – 10)
Subtraction	10	9	10	5*	9.2 (6 – 10)
Multiplication	10	9	10	3	7.4 (3 – 10)
Division	10	4*	10	4*	9.5 (7 – 10)
Higher order	4	1	4	4	3.0 (0 – 4)

* denotes abnormal performance

The patients' errors provide some support for this view. EK only made a substantial number of errors on the division problems. Two-thirds of these errors occurred because EK divided the components of a multi-digit number correctly but failed to add the products of these calculations (for example, she recorded half of 16 as '53', apparently because she knew that half of 10 was 5, and that half of 6 was 3. Similarly, she recorded half of 12 as '51'). Cappelletti et al. reported that IH, the SD patient in their study, had similar procedural problems with multi-digit multiplications (Cappelletti et al., 2001). EK's other errors were closer to the target number and consistent with the direction of the sequence. EK responded on every trial. In contrast, MK made frequent 'no response' errors on multiplication and division problems (accounting for 3/7 and 5/6 errors respectively), although she responded on every addition and subtraction trial. The majority of her other errors appeared to result from the use of the wrong mathematical procedure. 3/5 of her errors on the subtraction sequences occurred because she added the number that should have been subtracted. Similarly, 3/5 of her errors on the multiplication problems occurred because she added the previous number in the sequence.

3.6.4.4 Discussion

These results taken together provide evidence about which aspects of number knowledge remain intact in SD and which become degraded. The patients examined here were able to make accurate judgements about numerical magnitude (e.g., in comparison tasks), could place number words in the correct order, were able to translate between Arabic numerals and number words and could perform naming and word-picture matching with dot pictures. In contrast, their ability to perform calculations, particularly those requiring multiplication and division, was impaired. The patients showed a tendency to over-apply the addition procedure to other sums requiring different operations. Their knowledge of mathematical symbols was clearly degraded. They also performed more poorly on arithmetic problems that involved multi-digit operands, and some of their errors on these questions implied a specific impairment of the procedures required for multi-digit numbers. In addition, the patients made some errors on very easy multiplication problems (e.g., 2 x 3 = 8, 8 x 1 = 16), which may have been indicative of an impairment of arithmetic facts (McCloskey, 1992). Finally, although their ability to produce and understand number words appeared to be largely intact, EK and MK appeared to confuse number words like 'seventeen' and 'seventy' in several of these tasks (word-dot picture matching and number sequences tasks).

This pattern of competencies and weaknesses is consistent with models of numerical cognition that postulate a distinction between semantic/parietal representations of numerical magnitude and verbal representations of number words and facts (e.g., Dehaene, 1992; McCloskey et al., 1985; Noel & Seron, 1993). In McCloskey's model (McCloskey et al., 1985), for example, there is an amodal semantic representation of quantity that is recruited in all number tasks and several more peripheral mechanisms that are involved in the comprehension of number words/Arabic numbers, the retrieval of arithmetical facts and signs and the use of arithmetical procedures. Similarly, in Dehaene and Cohen's 'triple code' model (Dehaene, 1992), there is an analogue representation of numerical magnitude which is separable from the verbal code for number words and the visual code for Arabic numerals. Certain types of numerical processing (e.g., transcoding) are thought to draw heavily on the verbal and visual codes, whereas other tasks (e.g., magnitude comparison) are thought to be more independent of linguistic representations (but see Noel & Seron, 1997). Brain regions within the inferior parietal lobe are thought to underpin the analogue magnitude system in Dehaene's model (Dehaene & Cohen, 1995; Dehaene et al., 1998). Patients with SD should therefore have a good understanding of quantity, consistent with their intact performance on number comparison tasks. In contrast, the stable associations between verbal/Arabic numbers and representations of magnitude might become degraded in SD. Although the ability to comprehend number words relative to other words could be partially protected because numbers have meaningful referents that are frequently encountered in the world, there is some evidence that transcoding tasks involving single-digit numbers can become impaired in cases with very severe SD (Knott, 1998). Arithmetic facts are also stored in a verbal format in Dehaene's model, consistent with the patients' errors on easy multiplication problems.

It is also interesting to note that GT's number processing abilities were superior to those of EK and MK across a range of tasks. His particularly marked preservation of number knowledge may have been related to his extensive pre-morbid experience with number and calculation in his job as a college technician. Similarly, patient IH (Cappelletti et al., 2001) had exceptional number knowledge given his level of semantic impairment and had worked as a City banker. Therefore, pre-morbid experience of numbers may help to determine the extent to which they remain intact in this condition.

3.7 Reading aloud number and non-number words

If the patients' relatively good understanding of number helps to maintain the phonology of number words in ISR, a similar difference between number and non-number words might be expected to emerge in other apparently 'non-semantic' tasks requiring phonological production, for example, reading aloud. Some views about the translation from orthography to phonology suggest that semantic representations play an important role in reading aloud, especially for low frequency words with atypical spelling-to-sound correspondences (Plaut, McClelland, Seidenberg, & Patterson, 1996). SD patients make reading errors on such words, pronouncing them as if they had regular correspondences (PINT to rhyme with 'mint'): i.e., they demonstrate surface dyslexia (Graham, Hodges, & Patterson, 1994; Patterson & Hodges, 1992). In contrast, some models assume that semantics contributes little to reading aloud in skilled readers (Paap & Noel, 1991; Van Orden, 1987).

If semantic representations play an important role in reading aloud, the SD patients in this study might be expected to show preserved reading of irregular number words. The correct pronunciations of these items should receive more support from semantics if knowledge of the number domain is relatively preserved. Cappelletti et al. reported that patient IH, who had good comprehension of number words, did make fewer errors on reading aloud number words, compared with other categories of words, even when the words were matched for frequency and regularity (Butterworth et al., 2001; Cappelletti et al., 2002). EK, GT and MK were tested using Cappelletti et al.'s materials, in order to determine if they were also more accurate at reading number words compared with non-number words.

3.7.1 Method

The patients were asked to read 30 cardinal number words (the numbers from one to twenty, each tenth number - 'thirty', 'forty' etc, and the words hundred, thousand and million). They were also asked to read 22 ordinal number words (first, second, etc., up to twentieth, and then the items hundredth and thousandth), and 18 'ambiguous' number related words that also had non-numerical meanings (add, minus, share etc). These 70 number words were categorised as having regular and irregular spelling patterns and were compared with 70 non-number words matched on frequency, spelling regularity and length.

3.7.2 Results

The percentage of number and non-number words read aloud correctly by the three patients is shown in Table 3.10. Butterworth et al. (2001) reported that control participants were errorless on this task. The patients were relatively good at reading both regular and irregular number words, and made a larger number of errors on the non-number words, although the differences were subtle compared with the dramatic effect shown by IH. Chi-square tests were used to compare the balance of correct to incorrect items for the number and non-number words. The advantage for reading number words was statistically significant for both EK and GT (EK: $\chi^2(1) = 4.07, p < 0.05$; GT: $\chi^2(1) = 3.89, p < 0.05$). MK made a larger number of errors on both number and non-number words, and showed no significant differences between them ($\chi^2(1) < 1$). As anticipated, none of the patients showed an effect of regularity in their reading of number words ($\chi^2(1) < 1$). GT showed a marginally significant effect of regularity in his reading of non-number words (GT: $\chi^2(1) = 3.56, p = 0.059$), and the other two patients showed numerical advantages for regular over irregular non-number words that failed to reach significance (EK: $\chi^2(1) < 1$; MK: $\chi^2(1) = 1.13$, n.s.).

The patients made similar errors in reading aloud number and non-number words. Errors were classified as 'plausible' or 'implausible' pronunciations. Plausible pronunciations included regularisation errors (irregular words pronounced following the major correspondences described by Venezky (1970); for example, BASIC read as "bassik"). This category also included LARC errors (legitimate alternative reading of components: regular or irregular words pronounced according to Venezky's minor correspondences; for example, IMPLY read as "implee", ELEVEN read as "ell-even" to rhyme with "stephen"). The implausible category included errors that apparently resulted from letter or word confusions (e.g., FLAT read as "frat", PEACE read as "peach") and addition or deletion of letters/sounds (e.g., MONTHS read as "month", COMMENT read as "commentment"). A few 'mixed' errors appeared to result from a combination of plausible and implausible errors (e.g., NEGRO read as "nexro").

On the number words, EK made 5 implausible errors and no plausible errors. GT made a single plausible error and 2 implausible errors. MK made 4 plausible and 6 implausible errors. EK's implausible errors were all number words (SIXTEENTH pronounced as "seventeenth", EIGHTH pronounced as "eighteenth"). GT's implausible errors divided number words like SIXTEENTH into two number words, "six-tenth". In contrast, MK's implausible pronunciations of number words were, with one exception, not number words themselves (FOURTEENTH read as "portenth", NINETEENTH read as "dineteenth"). The ratios of plausible to implausible errors were similar for the non-number words (EK: 8:9, GT: 7:5, MK: 7:16).

Table 3.10: *Reading number and non-number words*

	EK	GT	MK
Cardinal number words (n = 30)			
Regular (n = 19)	100	100	84
Irregular (n = 11)	100	100	91
Ordinal number words (n = 22)			
Regular (n = 13)	77	77	77
Irregular (n = 9)	78	100	67
Ambiguous number words (n = 18)			
Regular (n = 14)	93	100	57
Irregular (n = 4)	75	100	50
Non–number words (n = 70)			
Regular (n = 46)	78	91	74
Irregular (n = 24)	71	71	54

Note: accuracy is expressed as a percentage of items presented

3.7.3 Discussion

EK and GT were better at reading numbers than non-number words matched for regularity, frequency and length, replicating the results of Butterworth et al. (Butterworth et al., 2001; Cappelletti et al., 2002). MK showed no difference between the number and non-number words, perhaps because her number comprehension was more impaired. GT showed an effect of regularity in his reading of non-number but not number words, which can be interpreted as indicating that stronger semantic representations for numbers enabled him to produce the correct phonology even for irregular items. The pattern of results was similar for EK and MK, but the regularity effect did not reach significance. In summary, the correct phonology was more likely to be produced for number than non-number words in both ISR and reading aloud, consistent with the notion that these items received stronger support from the semantic system.

3.8 General Discussion

This series of experiments examined immediate serial recall (ISR) of number and non-number words in patients with semantic dementia (SD), in order to investigate the quantity and quality of span performance for these two types of materials. For every patient, the recall of single-digit numbers was normal whereas the recall of non-number words was impaired relative to controls, and this number advantage extended to lower frequency multi-digit numbers and words. In every experiment, the patients' recall revealed a relatively normal pattern of omission, order and intrusion errors on the number words but an abnormally large number of phonological errors on the non-number words. The difference between numbers and non-number words remained substantial even when frequency, imageability, word length, set size and size of semantic category were matched across the two types of material.

The ISR differences between number and non-number words were reminiscent of those observed in SD patients for relatively well known vs. semantically degraded words (e.g. Chapter 2, Knott et al, 1997; Patterson et al, 1994), consistent with the suggestion that a comprehension difference might underpin the better ISR for numbers. Naming and word-picture matching tests with dot and face part pictures supported this hypothesis, as the patients were able to perform these comprehension tasks for the numbers but not the matched non-number words. A similar association between comprehension and ISR was also observed within the domain of numbers, in a comparison of multi-digit and single digit number words. Moreover, the patient with the largest comprehension impairment for the non-number words also exhibited the poorest ISR for these words.

The finding of selectively preserved ISR for number words raises the question of the extent to which number comprehension is preserved in SD. Cappelletti et al.'s patient IH (Butterworth et al., 2001; Cappelletti et al., 2001; Cappelletti et al., 2002) appeared to have good number comprehension despite his severe semantic impairments in other domains, as he could translate between Arabic numerals and English number words, make accurate judgements about numerical order and magnitude and read and spell number words more accurately than other words. The present research largely replicated these findings. The patients investigated here were able to perform number comparison and ordering tasks, suggesting that their understanding of numerical magnitude was relatively good. They were also able to translate between Arabic numerals/dot pictures and number words and showed an accuracy difference in reading aloud number and non-number words.

Some aspects of the patients' mathematical abilities were clearly impaired, however. In line with previous reports (Cappelletti et al., 2001), the patients' calculation abilities were compromised by impaired procedural knowledge, particularly of multiplication and division. They performed more poorly on arithmetic problems that involved multi-digit operands, and some of their errors on these questions implied that their knowledge of the procedures required for such problems was impaired. Their knowledge of numerical symbols was also clearly degraded. In addition, the patients made some surprising errors on very easy multiplication problems (e.g., 2 x 3 = 8), which may have been indicative of an impairment of arithmetic facts (McCloskey, 1992). These deficits were most striking for patient MK, who had the greatest impairment of semantic memory. In addition, it appears that pre-morbid experience with numbers might affect the degree to which they remain intact in this condition. GT was exceptionally good at number tasks (better than EK who had less severe semantic impairments), and he had considerable experience with numbers in his job as a college technician. Patient IH (Cappelletti et al., 2001) also had exceptional number knowledge that was out of step with his severe semantic impairments, and extensive pre-morbid experience with numbers in his occupation as a City banker.

The patients' performance across a number of tasks was generally better for high frequency single-digit numbers than for lower frequency multi-digit numbers. MK, the patient with the most severe semantic impairment in this study, provided a particularly striking example of this. She was able to generate the next number for 'higher order' sequences in which the amount to be added or subtracted was changed by a constant amount each time, suggesting that she had an excellent understanding of these numbers, all of which were under 25. However, she showed a complete inability to place multi-digit lower frequency numbers in the correct order, even though this task did not involve calculation and arguably required a less subtle understanding of numerical magnitude. Concepts are thought to degrade gradually rather than in an all or none fashion in SD (Hodges, Graham, & Patterson, 1995) and the same may be true of number words. The patients in this study retained a good understanding of single-digit numbers but their knowledge of less frequent multi-digit numbers was apparently somewhat impaired, consistent with the greater vulnerability of low frequency words and concepts in this condition (Funnell, 1995). However, single-digit number words also refer to concepts that are easy to manipulate and acquired at a young age; factors which may have contributed to their better comprehension and recall.

Why is number knowledge relatively preserved in SD? Cappelletti et al. suggested that straightforward explanations, like the high frequency and orderliness of numbers, could not account for IH's superior understanding of number because non-number words that shared these characteristics were not preserved (Cappelletti et al., 2001). Similarly, this chapter demonstrates that the selective preservation of ISR for numbers survives matching for frequency, imageability, word length, set size and open or closed semantic category. Another possibility is that numbers are preserved because they are important in everyday life; for example, they are essential for shopping and to use the telephone. Patients may retain an understanding of words and concepts that they encounter every day because of their preserved ability to form new episodic memories (Snowden, Griffiths, & Neary, 1994; Snowden, Griffiths, & Neary, 1996). Alternatively, number knowledge may be relatively preserved in SD because the cortical atrophy associated with this condition typically spares the brain regions thought to be crucial for the representation of numerical magnitude. The participants in this study had marked atrophy of the inferolateral temporal lobe, in common with other SD patients (Hodges, Patterson, Oxbury, & Funnell, 1992; Snowden, Goulding, & Neary, 1989). In contrast, the inferior parietal lobe appears to be critical for numerical understanding. This brain region is damaged in patients with acalculia (e.g., Cipolotti et al., 1991; Dehaene & Cohen, 1997; Delazer & Benke, 1997; Takarama et al., 1994; Warrington, 1982) and is apparently intact in patients who have preserved numerical abilities despite global cognitive decline

(Remond-Besuchet, Noel, Thioux, Brun, & Apse, 1999). The inferior parietal lobe is also activated by number processing tasks in brain imaging studies (Dehaene et al., 1999; Dehaene et al., 1996; Pinel et al., 2001; Stanescu-Cosson et al., 2000).

Some aspects of numerical understanding appear to be related to particular linguistic abilities. In number transcoding tasks, for example, patients with Broca's aphasia make predominantly syntactic errors (e.g., writing 125 as "10025"), whereas patients with Wernicke's aphasia make a larger number of lexical errors (Deloche & Seron, 1982b). These impairments are supposedly underpinned by specific difficulties in processing syntactic structures and accessing lexical forms respectively. On the other hand, some aspects of number knowledge appear to be independent of language. Patients with severe acalculia can nevertheless have intact language (Cipolotti et al., 1991; Delazer & Benke, 1997) and conversely, patients with severe anomia can have intact numerical skills (Cappelletti et al., 2001; Rossor, Warrington, & Cipolotti, 1995). Various models of numerical cognition have proposed a distinction between representations of number that use a linguistic code (e.g., 'lexical' representations of English number words), and non-verbal representations of numerical magnitude (Dehaene, 1992; McCloskey et al., 1985; Noel & Seron, 1993). Numerical processes that draw heavily on magnitude representations might be expected to remain relatively intact in SD, whereas, in contrast, tasks that involve linguistic number representations might be expected to degrade over the course of the condition. If this conjecture is correct, it might be expected that ISR and comprehension of number words would be initially protected in SD because these words refer to meaningful parietal lobe representations of numerical magnitude. Eventually, however, transcoding and ISR tasks should become impaired even for single-digit numbers. There is some evidence to support this view. Knott (1998) reported that patient FM, who had very severe SD, was impaired at naming and word-picture matching with the numerals 1-9 and also made phonological errors during digit span. This association between the understanding and recall of single-digit numbers lends further support to the view that semantics plays a crucial role in verbal STM.

In summary, the work presented in this chapter explored the issue of why SD patients show intact digit span in the context of poor ISR for non-number words. Four patients showed better immediate recall of number than matched non-number words and made fewer phonological errors on these items. The patients also showed better comprehension of the number words. This finding is consistent with the view that the recall difference between number and non-number words is an instance of the ISR difference between known and semantically degraded words. However, it could be argued that the recall of number words was relatively intact in these patients because numbers are semantically impoverished – ISR for numbers, particularly single-digit numbers, may be less dependent on the semantic system than the recall of non-number words in healthy individuals. A similar argument is adopted in Chapter 4 to account for the patients' relatively intact recall of nonwords. Whatever the merits of this line of reasoning, this chapter has provided evidence for a category specific advantage for numbers in ISR that corresponds with a similar advantage in picture naming and word-picture matching tasks. This association at the very least provides support for the claim that phonological coherence in ISR is linked to language production and comprehension.

4 Evidence for intact phonology in semantic dementia

4.1 Introduction

Semantic dementia patients are generally considered to have intact phonology, because they rarely, if ever, make phonological errors in spontaneous speech or in naming to confrontation. Likewise, their immediate repetition of single words is excellent (Knott, Patterson, & Hodges, 1997). In contrast, as documented in Chapters 2 and 3 and by several previous studies (Knott et al., 1997; Knott, Patterson, & Hodges, 2000; Patterson, Graham, & Hodges, 1994), their ISR performance is characterised by phonological breakdown, particularly for semantically degraded words. This finding is consistent with the suggestion that semantic memory makes an important contribution to the coherence of phonological representations in STM tasks (e.g., Patterson et al., 1994; N. Martin & Saffran, 1997). As discussed in Chapter 2, however, several studies have failed to find the expected superior recall of known over degraded words (Funnell, 1996; Lambon Ralph & Howard, 2000; McCarthy & Warrington, 1987, 2001; Warrington, 1975), challenging this view. These findings have suggested to some researchers that additional phonological or lexical impairments, independent of the patients' primary semantic deficits, might account for the poor ISR observed in some cases (Knott et al., 1997; McCarthy & Warrington, 2001). It is important to note, on the other hand, that the SD patients in all of these studies made frequent phonological errors in ISR even if they did not show a recall advantage for known words, suggesting a strong association between semantic impairment and the integrity of representations in phonological STM.

This chapter aims to address the issue of whether additional phonological or lexical deficits should be invoked to account for the ISR impairments of SD patients. The question of whether phonology is intact in SD remains largely neglected, despite its importance. Certainly, SD patients very rarely make phonological errors in spontaneous speech or picture naming (Patterson & Hodges, 2000; Snowden, Griffiths, & Neary, 1994), suggesting that they do not have problems with assembling the phonological elements of single words. Intact performance has been reported for a few patients on phonological awareness tasks like minimal pair discrimination (detecting that two phonologically similar words – e.g. cup/cut – are different), phoneme segmentation tasks (requiring phonemes to be added to or deleted from words) and rhyme generation (Knott et al., 1997, 2000). In addition, SD patients have normal digit span abilities (see Chapter 3) and digit span remains relatively stable in the face of marked semantic decline (Knott et al., 2000). Moreover, a few patients have been shown to exhibit normal effects of phonological similarity in ISR (Chapter 3; Knott et al., 2000; McCarthy & Warrington, 2001), suggesting that verbal STM in SD relies heavily on a phonological code, as it does in healthy individuals. Normal effects of word length in ISR have also been demonstrated (Knott et al., 1997, 2000), although not for every patient tested (McCarthy & Warrington, 2001).

Some aspects of the ISR performance of SD patients, however, raise doubts about the integrity of the phonological system. First, SD patients have been shown to exhibit little effect of recency in their serial position curves (see Chapter 2 and Knott et al., 1997), a pattern which is associated with phonological impairments in aphasic populations (N. Martin & Saffran, 1997). Secondly, ISR in SD appears to be strongly influenced by word frequency and imageability (Knott et al., 1997, 2000; McCarthy & Warrington, 2001), even though lexical and semantic effects in ISR are generally reduced in patients who have difficulty retaining semantic information (R. C. Martin & Lesch, 1996; R. C. Martin, Shelton, & Yaffee, 1994) and enhanced in patients with phonological impairments (N. Martin & Saffran, 1997). Thirdly, Knott et al. (1997) demonstrated that patient AB exhibited rapid phonological decay of single words in a delayed repetition paradigm, when the delay was filled with counting. Lastly, Knott et al. (1997) obtained some evidence of a nonword repetition deficit in AB. This patient showed a reduction of the normal lexicality effect in ISR, suggesting that his recall was more impaired for words than for nonwords. AB was, however, only able to repeat 28/40 multisyllabic nonwords (Children's Test of Nonword Repetition; Gathercole, Willis, Baddeley, & Emslie, 1994), a score which would be impaired for nine year-old children. Similarly, his recall of single syllable nonword sequences was right at the bottom of the normal range and was quite possibly impaired relative to his premorbid abilities.

None of these findings necessarily points to additional phonological or lexical deficits in SD. The lack of a recency effect could occur because semantic deficits impact most heavily on the phonological integrity of items at the end of lists. The phonological representations of these words may become particularly noisy in the absence of semantic support because they must be maintained for longer and during the production of other potentially interfering items. Hulme et al. (1997) found larger effects of word frequency in the recency portion of the serial position curve for normal participants, in line with this suggestion. The delayed single word repetition results of Knott et al. (1997) could be interpreted in the same way as the ISR differences between known and degraded words, namely, as a loss of semantic binding. The items were not selected according to patient AB's knowledge of them but they were probably partially semantically degraded. Consequently, AB may have been unable to sustain their phonology during the distraction task because of this reduced semantic support. In addition, strong effects of frequency may arise in SD because lower frequency words generally degrade earlier in this condition than more frequent items (Funnell, 1995).

Although nonword recall is thought to be a purer measure of phonological STM than word recall because it involves unfamiliar phonological forms (Gathercole et al., 1994), word knowledge does appear to contribute to nonword ISR. Nonword recall is affected by 'wordlikeness' that is, the degree to which nonwords are rated as being similar to words (Gathercole, 1995; Gathercole, Willis, Emslie, & Baddeley, 1991). This 'wordlikeness' effect may be underpinned by both sublexical factors like phonotactic frequency (Gathercole, Frankish, Pickering, & Peaker, 1999; Gathercole & Martin, 1996) and by lexical factors such as the number of real-word phonological neighbours a nonword has. Recent research suggests that the lexical contribution to nonword recall may be stronger than the sublexical one (Bailey & Hahn, 2001; Roodenrys & Hinton, 2002). By extension, nonword repetition impairments in SD could be underpinned by deficits in word knowledge rather than by impairments of phonology per se. One possibility, suggested by Knott et al. (1997), is that abnormal nonword repetition results from lexical impairments that are independent of these patients' primary semantic deficits. This suggestion concurs with the common sense view that semantic factors do not impact on nonword recall, as these stimuli are basically meaningless. However, the PDP framework (Patterson et al., 1994; Plaut & Kello, 1999) might predict an effect of semantic impairment on nonword repetition. If phonology is considered to emerge from the interactions between speech input, articulation and semantics (Plaut & Kello, 1999), then phonological space is established in a semantic context. It follows from this that semantic degradation will have a considerable impact on the operation of the phonological system, as it will change the nature of the phonological space.

The work presented in this chapter examined the performance of six SD patients and matched controls on a range of phonological processing and STM measures in order to evaluate the claim that the phonological system is unimpaired in this condition. The influence of phonological similarity, word length and lexicality on ISR performance was compared for the two groups and the patients' recall of relatively well-known and semantically degraded words was examined to determine whether a recall advantage for known words could emerge in the absence of additional phonological deficits. This study also examined the patients' recall of nonwords constructed from known and degraded words in order to explore the contention that semantic knowledge contributes to nonword recall.

4.2 Case descriptions

This work (largely carried out from June to December 2002) examined six SD patients, who are described below in order of severity. As all the cases exhibited similar patterns of deficits, the details given for the first patient can be assumed to apply to every case unless otherwise stated. A summary of the background neuropsychological assessment is shown in Table 4.1. SJ, the least impaired patient, was a 60-year-old right-handed woman who had been experiencing worsening word-finding difficulties for approximately three years. She left school aged 16 and was working part-time as an antiques dealer at the time of the study. Her neuropsychological profile was dominated by a mild to moderate impairment of semantic memory. An MRI scan from 2001 showed bilateral temporal lobe atrophy that was more marked in the left hemisphere. She was impaired on tests requiring comprehension of words and pictures; for example, word-picture matching and the Pyramids and Palm Trees test (Howard & Patterson, 1992). She was anomic in spontaneous speech, word fluency tasks and confrontational picture naming. Her naming errors were predominantly omissions and semantic paraphasias. In common with other SD patients, she produced surface dyslexic errors in reading aloud and surface dysgraphic errors in spelling tasks. In contrast to her semantic difficulties, she was well oriented in time and place, had excellent episodic memory for recent events, and had no difficulty in remembering appointments. She performed normally on tests of visual-spatial processing from the Visual Object and Space Perception battery (VOSP, Warrington & James, 1991), and she was able to produce a good immediate copy of the Rey complex figure (Lezak, 1976). Her non-verbal reasoning on the Raven's Coloured Progressive Matrices test (Raven, 1962) was normal. Her speech was fluent and syntactically well formed despite her anomia. She had intact single word phonology and she did not make phonological errors in her spontaneous speech or picture naming. She had normal spatial STM as assessed by the Corsi block tapping task, and normal verbal STM as measured by forwards and backwards digit span (Wechsler, 1987). Her word span performance, however, was characterised by frequent phonological errors similar to those described by Patterson et al. (1994).

BS was a 67-year-old right-handed man who left school aged 16 and had previously worked as a bookseller. He had been experiencing a gradual decline in semantic memory for around four years. He first reported difficulties in recognising the faces of acquaintances but by the time of the study he also experienced frequent word-finding difficulties. An MRI scan from 2002 showed marked bilateral temporal lobe atrophy. His cognitive profile was similar to the description of SJ above although his semantic impairments were a little more severe.

EK, who is also described in Chapter 2, was 60 years old at the time of the study. Her semantic performance had deteriorated to some extent between the assessments presented in Table 4.1 and those described in Table 2.2 (a period of approximately one year). Her word-picture matching performance had dropped from 46/64 to 39/64, for example. In contrast, she still performed normally on tests of visual-spatial processing and produced a good copy of the Rey figure. Her non-verbal reasoning skills were unchanged. Her forwards digit span did not decline over this period and she still showed normal single word phonology and good episodic memory for recent events. Her history and personal details are described in section 2.2.

KI was a 65-year-old right-handed man, who left school age 14 and had previously worked in heavy engineering. He had a four-year history of worsening semantic impairments. He was severely impaired on both verbal and pictorial tests of semantic memory although, unusually, he exhibited greater deficits on pictorial tests. He had considerable difficulties recognising objects and faces and he was severely anomic in spontaneous speech, fluency tasks and confrontational picture naming. In contrast, he had good visual spatial processing and episodic memory for recent events. He showed some weakness, however, in a test of non-verbal reasoning. He also exhibited behavioural changes, including disinhibition, which would be consistent with the disease process affecting basal frontal as well as temporal regions (Snowden, Neary, & Mann, 1996).

Table 4.1: *Background neuropsychological scores (2002)*

Test	Max	SJ	BS	EK	KI	JT	GT	Controls M	Controls SD
MMSE[1]	30	23*	25	26	23*	25	22*	> 24[a]	-
Coloured Progressive Matrices[2]	36	34	30	33	21*	36	35	-	-
Digit span: forwards[3]	-	5	8	7	8	8	7	6.8[b]	0.9[b]
Digit span: backwards[3]	-	3	4	4	5	4	4	4.7[b]	1.2[b]
Spatial span: forwards[4]	-	6	NT	6	6	5	5	5 - 6[c]	-
Naming	64	30*	29*	18*	15*	6*	11*	62.3[b]	1.6[b]
Word-picture matching	64	59*	40*	39*	36*	34*	27*	63.7[b]	0.5[b]
PPT: Pictures[5]	52	48*	33*	30*	31*	35*	32*	51.1[b]	1.1[b]
PPT: Words[5]	52	42*	35*	35*	35*	31*	27*	51.2[b]	1.4[b]
Category fluency (8 categories)	-	31*	45*	27*	27*	9*	11*	113.9[d]	12.3[d]
Letter Fluency (F, A, S)	-	23*	33*	27*	17*	17*	14*	44.2[b]	11.2[b]
Rey figure immediate copy[6]	36	33	33	36	35	34	33	34.0[d]	2.9[d]
VOSP: incomplete letters[7]	20	20	19	20	8*	18	17	19.2[b]	0.8[b]
VOSP: dot counting[7]	10	10	10	9	10	9	10	9.9[b]	0.3[b]
VOSP: position discrimination[7]	20	20	19	20	19	18	20	19.8[b]	0.6[b]
VOSP: cube analysis[7]	10	10	10	10	10	10	9	9.7[b]	2.5[b]

* denotes abnormal performance (i.e., more than two standard deviations below the control mean); NT denotes not tested. Figures show number of items correct.

[1] Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975)

[2] Raven's Coloured Progressive Matrices (Raven, 1962)

[3] Weschler Memory Scale - Revised (Wechsler, 1987)

[4] Weschler Memory Scale – III (Wechsler, 1997)

[5] Pyramids and Palm Trees Test (Howard & Patterson, 1992)

[6] Rey figure taken from Lezak (1976)

[7] Visual Object and Space Perceptual Battery (Warrington & James, 1991)

[a] Cutoff for normal performance

[b] Control data from Bozeat et al. (2002)

[c] Normal range for age matched participants

[d] Control data from Hodges and Patterson (1995)

JT, a 66-year-old right-handed male who left school at 16, was running a small farming business at the time of the study. He had been experiencing worsening word-finding difficulties for four years. An MRI scan from 2002 showed significant temporal lobe atrophy that was considerably more marked on the left side. He was severely impaired on a range of pictorial and verbal tests of semantic memory and his picture naming performance was approaching floor. However, his single word phonology, visual-spatial skills, non-verbal reasoning abilities and memory for recent events were largely intact.

GT, who is also described in Chapter 2, was 71 years old at the time of the study. His semantic performance had deteriorated between the assessments presented in Tables 4.1 and 2.2; his word-picture matching performance had dropped from 32/64 to 27/64, for example. In contrast, he still performed normally on tests of visual-spatial processing and produced a good copy of the Rey figure. His digit span and non-verbal reasoning skills were unchanged. His single word phonology and episodic memory for recent events remained largely intact. Section 2.2 provides further information about GT's case history.

4.3 Phonological processing abilities

4.3.1 Method

The six patients were tested on a variety of tasks thought to tap phonological processing skills. First, they were given a phoneme segmentation task (from Patterson & Marcel, 1992) that required phonemes to be deleted from and added to the beginning of words and nonwords. In the phoneme addition task, the examiner read a word like 'old' and asked the patient to join the sound 'g' onto the beginning, to make 'gold'. In the phoneme deletion task, the examiner presented 'gold' and asked the patient to take away the first sound, leaving 'old'. There were 48 trials of each type, blocked using an ABBA design. Within each block of 24 trials, there were six trials in which the patient was given a word and had to produce a second word in response ('old' into 'gold'). In a further six trials, the stimulus was a word and the patient had to produce a nonword ('shave' into 'ayve'). In six more trials, the stimulus was a nonword and the patient had to produce a word ('pice' into 'ice'), and in the final six trials, the stimulus and target were both nonwords ('vafe' into 'aff'). These four conditions were presented in a mixed fashion. Patients practised the task until it was clear that they understood it.

The patients were also given the minimal pairs tests from the PALPA battery (Kay, Lesser, & Coltheart, 1992). They were tested on two parallel tasks involving CVC (consonant vowel consonant) words and nonwords. Both tasks required the patients to detect that two phonologically similar items (e.g. 'gut' and 'cut') were different. There were 36 'no change' trials and 36 'change' trials in each test. Stimulus pairs were minimally different according to position (initial phoneme, final phoneme or metathetic differences in which the initial and final phonemes were exchanged) and type of change (voicing, manner or place of articulation). There were equal numbers of each type of change.

In a further test of their phonological processing abilities, the patients were given rhyme judgement and production tasks (from Patterson & Marcel, 1992). In the rhyme judgement task, the patients were presented with pairs of words like 'fall-call' and were asked to decide if the pairs rhymed. There were 48 pairs in total, comprising an equal number of rhyming and non-rhyming pairs. The words in the non-rhyming pairs were either phonologically similar or dissimilar. In the rhyme production task, 24 words were read aloud, and for each one the patient was asked to think of a word that rhymed with it. These rhyme tests had to be abandoned for GT, as he was unable to understand the test instructions, despite being given numerous examples of rhyming and non-rhyming words.

4.3.2 Results

Table 4.2 gives the patients' scores on these tests of phonological processing. The patients with the mildest semantic impairments, SJ and BS, performed normally on the minimal pairs and phoneme segmentation tasks but showed some weakness on the rhyme judgement and production tasks. They were able to think of rhyming words for the majority of items in the rhyme production test but they deviated from the normal pattern by sometimes producing rhyming nonwords. In the rhyme judgement task, they showed a tendency to accept phonologically similar words as rhymes, perhaps because they did not fully comprehend the notion of rhyme.

EK and KI, who were rather more severely semantically impaired, also performed normally on the minimal pairs task but exhibited some impairment of phoneme segmentation. They tended to make phonological errors on this task (e.g., 'n' added to 'oath' produced as 'note' and not 'noath'). In many of these errors, the phoneme of interest was added or subtracted correctly but errors occurred in other parts of the target. As the memory load in the segmentation task was considerable, these errors may have reflected the impact of the patients' semantic deficits on the coherence of items in phonological STM. In line with this suggestion, the patients were more impaired on the addition than the subtraction version of the task, perhaps because the addition version required simultaneous maintenance of both the item and the phoneme to be added and therefore made greater STM demands. Alternatively, the addition task may have been more demanding because it required subtraction of the neutral schwa sound from the phoneme to be added as well as phoneme addition. It is also worth noting that 6/15 of EK's errors involved the 'sh' phoneme. She appeared to be using a letter-spelling strategy (add or take away initial letter) that failed for this multi-letter phoneme and resulted in errors in the subtraction task like 'hade' for 'shade', instead of 'aid'. JT, who had a more severe semantic impairment than EK and KI, did not show this weakness in the phonological segmentation task but was more substantially impaired in the rhyme production task.

Table 4.2: *Performance on tests of phonological processing*

Test	Max	Control mean (SD)	SJ	BS	EK	KI	JT	GT
Phoneme segmentation: addition	48	[a]	45	48	37*	38*	44	32*
Phoneme segmentation: subtraction	48	[a]	47	48	44	45	46	34*
Minimal pairs: words (same) [1]	36	35.54 (0.78)	36	35	36	36	36	35
Minimal pairs: words (different) [1]	36	34.54 (2.58)	36	36	35	33	34	29
Minimal pairs: nonwords (same) [1]	36	35.79 (0.56)	34	35	36	36	36	35
Minimal pairs: nonwords (different) [1]	36	35.09 (2.34)	36	36	35	35	34	25*
Rhyme judgement: total correct	48	[a]	43*	46	42*	40*	46	-
Rhyme judgement: non-rhyming phonologically similar pairs	12	[a]	8	10	6	7	12	-
Rhyme production: correct real-word responses	24	[a]	21*	23	20*	20*	17*	-
Rhyme production: all rhyming responses	24	[a]	22	24	23	23	20*	-

* denotes abnormal performance. Figures show number of items correct.

[1] from the PALPA battery (Kay et al., 1992)

[a] controls perform at ceiling on these tasks (Bird, Lambon Ralph, Seidenberg, McClelland, & Patterson, 2003)

The patient with the most severe semantic deficits in this study, GT, showed considerable weakness in all of these tests of phonological processing. First, he was the only patient to show a deficit on the minimal pairs tests. His 'same' judgements were normal, but he was poor at discerning when two items were different, particularly for nonwords. His performance on word stimuli was right at the bottom of the range of scores expected for healthy participants and his performance on nonwords was more substantially outside the normal range. Therefore, although the difference between words and non-words did not reach significance ($\chi^2(1) < 1$), there was some suggestion of a lexicality effect in his performance on this task. Combining across words and nonwords, GT's performance was not affected by whether the minimal difference occurred in the initial or final positions of pairs or in metathetically related pairs: he detected 20/24 differences in the initial position, 16/24 in the final position and 18/24 metathetic differences. His performance was substantially affected by whether the pairs were minimally different according to voice, manner or place of articulation ($\chi^2(2) = 40.44$, $p < 0.0001$). He detected 24/24 differences in voicing, 23/24 differences in manner but only 7/24 differences in place of articulation, possibly because of his slight hearing impairment. GT also exhibited the most substantial impairment on the phonological segmentation test observed in this group of patients. When the results from the addition and subtraction versions of the test were combined, he showed a significant effect of lexicality ($\chi^2(2) = 9.01$, $p < 0.05$). He was better able to perform the task when both the stimulus and target were real words (19/24) than when they were both nonwords (13/24). He showed an intermediate level of performance when the stimulus was a word and the target a nonword (16/24), and when the stimulus was a nonword and the target a word (18/24). These two categories were combined in the chi-square analysis.

4.3.3 Discussion

The patients' performance was not entirely intact on these tasks of phonological processing, although their poor performance might have resulted from their primary semantic impairments and not from any additional phonological deficits. There appeared to be an association between semantic impairment and poor phonological processing, as the most semantically impaired patients in this study also showed the poorest performance on these tasks. There are, however, two plausible explanations of this relationship. First, the temporal lobe atrophy that characterises SD may encroach on phonological as well as semantic areas as the condition progresses. Alternatively, if semantics plays a major role in the coherence of phonological representations, as suggested by the semantic binding hypothesis (Patterson et al., 1994), then the degree of semantic impairment should impact on performance in these tasks even in the absence of an additional phonological impairment. The phoneme segmentation, minimal pairs and rhyme judgement tasks all required accurate phonological representations to be maintained. In addition, GT showed an influence of lexicality in his segmentation and minimal pairs performance. Although this result suggests that these tasks may be influenced by the integrity of the semantic system, it is also possible that GT showed an effect of lexicality precisely because he had a phonological processing deficit. Lexical-semantic effects may become exaggerated when the phonological system is impaired. Clearly, the patients' performance on these phonological processing tasks cannot be interpreted unambiguously. For this reason, the effects of phonological similarity and word length on the patients' ISR were examined, as these well-documented effects are considered to be the hallmarks of normal phonological STM and articulatory rehearsal respectively (e.g., Vallar & Papagno, 2002).

4.4 Phonological similarity effects in immediate serial recall

Verbal STM in healthy individuals is poorer for similar sounding items compared with phonologically more distinct items. This 'phonological similarity effect' is usually taken as evidence for phonological coding in verbal STM (Baddeley, 1966; Conrad, 1964; Conrad & Hull, 1964). Therefore, a normal effect of phonological similarity in patients with SD would be consistent with intact phonological coding in verbal STM.

4.4.1 Method

The patients and twelve control participants matched for age and educational level were asked to recall letters from the phonologically similar set E, C, T, P, V, B, G, D or the phonologically dissimilar set W, S, Q, Y, R, J, F, L, following the method outlined in Chapter 3. The phonologically similar and dissimilar letters were blocked using an ABBA design. They were read aloud in lists of four and six items at a rate of one item per second for immediate serial recall.

4.4.2 Results

4.4.2.1 Recall accuracy

Table 4.3 gives the number of phonologically similar and dissimilar items and lists that were recalled correctly by the patients and controls. The performance of EK and GT on this task was discussed briefly in Chapter 3 but is reproduced here to allow a comparison with the other patients. The item data were first examined using an analysis of variance (ANOVA) comprising two within-subject factors (phonological similarity and list length) and one between-subject factor (group: patients versus controls). There were significant effects of phonological similarity ($F(1, 16) = 45.93, p < 0.0001$) and list length ($F(1, 16) = 109.66, p < 0.0001$). The main effect of group was not significant ($F(1, 16) < 1$), suggesting that letter recall was not greatly impaired in the patients. There was also no evidence of a phonological similarity by group interaction ($F(1, 16) = 1.67$, n.s.), suggesting that the magnitude of the phonological similarity effect did not differ very substantially for the patients and controls. No other interactions reached significance.

Table 4.3: *Recall of phonologically similar and dissimilar letters*

		List length	SJ	BS	EK	KI	JT	GT	Control mean (range)
	Dissimilar	4	97.5	97.5	100.0	95.0	90.0	90.0	96.5 (85 – 100)
Item recall	Similar	4	70.0	87.5	90.0	95.0	92.5	85.0	82.7 (63 – 98)
	Difference	4	27.5	10.0	10.0	0.0	-2.5*	5.0	13.8 (0 – 35)
	Dissimilar	6	73.3	85.0	86.7	70.0*	71.7	81.7	83.8 (72 – 97)
Item recall	Similar	6	65.0	63.3	70.0	61.7	71.7	75.0	70.1 (53 – 82)
	Difference	6	8.3	21.7	16.7	8.3	0.0*	6.7	13.6 (5 – 23)
	Dissimilar	4	90	90	100	80	60	70	88.3 (50 – 100)
List recall	Similar	4	10	60	60	80	70	40	50.0 (10 – 80)
	Difference	4	80	30	40	0	-10*	30	38.3 (0 – 90)
	Dissimilar	6	10	40	50	0*	30	20	40.0 (10 – 80)
List recall	Similar	6	0	10	0	0	0	20	8.3 (0 – 40)
	Difference	6	10	30	50	0*	30	0*	31.7 (10 – 50)

* denotes abnormal performance. Figures show percentage of item and lists correct.

Considering each patient individually, the number of letters recalled fell within the normal range for every patient tested except KI, who showed a mild weakness in some conditions (see Table 4.3). KI was the only patient in this group who was impaired at naming incomplete letters in the VOSP test battery (Warrington & James, 1991: see Table 4.1), suggesting that his immediate recall deficits for letters may have been underpinned by degraded knowledge of letters. SJ showed a significant phonological similarity effect ($t(37) = 3.28, p < 0.01$), as did BS ($t(32) = 2.93, p < 0.01$) and EK ($t(35) = 3.33, p < 0.01$). In contrast, the three more severely impaired patients did not recall the phonologically similar letters more poorly than the dissimilar letters (KI and JT: $t(37) < 1$; GT: $t(35) < 1$), suggesting that the phonological similarity effect may have been reduced in size in these patients. It is important to note however, that the magnitude of the phonological similarity effect was very variable in the control participants (in line with the results of Logie, Della Sala, Laiacona, Chambers & Wynn, 1996), and while every control showed a numeric advantage for dissimilar items, this did not always reach significance ($t(25\text{-}38) = 4.63, p < 0.0001$ to $t < 1$). These analyses combined across list length, although the pattern of results did not change if the longer lists were considered separately. The size of the phonological similarity effect was within the normal range for every patient except JT, who showed a slight numerical advantage for phonologically similar lists on the shorter list length. None of the controls showed a difference in this direction, although their recall was constrained by ceiling effects. On the longer lists, the size of JT's phonological similarity effect was again outside the normal range for item recall, but was more normal for list recall.

4.4.2.2 Error analysis

The errors made by the patients and controls on phonologically similar and dissimilar letters are shown in Table 4.4. Omission errors occurred if fewer items were recalled than were presented. Order errors were targets produced in the wrong place in the sequence. Repetition errors were targets recalled more than once in a list. Within-set intrusions occurred when letters from the set were recalled in the wrong list. Outside-set intrusions occurred when participants produced letters that did not form part of the experimental set.

Phonological errors were responses that did not fall into any of the previous categories and contained at least half of the phonemes in a target word. The controls made a larger number of order errors ($t(11) = 3.01$, $p < 0.05$), repetition errors ($t(11) = 3.29$, $p < 0.01$) and within-set intrusion errors ($t(11) = 4.27$, $p < 0.01$) in their recall of phonologically similar compared with dissimilar letters, combining across list length. The patients made a similar pattern of errors, although the number of order, repetition and outside set intrusions exceeded the normal range for some patients. As a group, the patients made a significantly larger number of within-set intrusions in their recall of phonologically similar compared with dissimilar letters ($t(5) = 3.25$, $p < 0.05$). The difference between phonologically similar and dissimilar letters also approached significance for repetition errors ($t(5) = 2.19$, $p = 0.08$). Neither the patients nor the controls made substantial numbers of phonological errors in letter recall. In line with the view that phonological similarity had a comparable effect on the errors made by patients and controls, there was no interaction between participant group, phonological similarity and error type in an ANOVA ($F(5, 80) < 1$). The interaction between phonological similarity and error type reached significance ($F(5, 80) = 8.63$, $p < 0.0001$), presumably reflecting the fact that, for both the patients and controls, the effect of phonological similarity was largely underpinned by an increase in repetition and within-set intrusion errors. In addition, there was an interaction between participant group and error type ($F(5, 80) = 4.19$, $p < 0.01$) that most likely resulted from the greater prevalence of repetition errors in the patients.

4.4.3 Discussion

On the whole, the patients showed effects of phonological similarity that were within the normal range. Further discussion of these findings will be postponed until after the next experiment, which considers the effect of word length on ISR.

Table 4.4: *Errors made by patients and controls on phonologically similar and dissimilar letters*

		Patients						Controls	
		SJ	BS	EK	KI	JT	GT	Mean	Max
Dissimilar	Omission	1	2	0	2	2	1	1.8	5
	Order	6	3	5	5	10	6	4.5	10
	Repetition	5*	4	1	13*	7*	3	1.4	4
	Within set intrusion	0	1	2	1	5	5	3.0	6
	Outside set intrusion	6*	0	1	1	0	1	1.0	3
	Phonological	1	0	0	0	0	0	0.0	0
Similar	Omission	0	5	2	0	1	3	3.1	11
	Order	13*	8	8	5	5	2	7.7	11
	Repetition	7	6	8*	21*	6	4	3.5	7
	Within set intrusion	13	7	7	3	6	12	8.4	17
	Outside set intrusion	0	0	0	0	0	0	0.4	5
	Phonological	0	0	0	0	0	0	0.0	0

* denotes abnormal performance. Figures indicate error counts summed across both list lengths.

4.5 Word length effects in immediate serial recall

Healthy participants are able to recall a larger number of short than long words in verbal STM. This 'word length effect' has been variously attributed to a time-based rehearsal process used to reinstate the phonological trace as it decays (Baddeley, Lewis, & Vallar, 1984; Baddeley, Thomson, & Buchanan, 1975), to decay during speech output delays (Cowan et al., 1992) and to phonological complexity, as longer words are generally also more complex (Service, 1998). Although there is still controversy about the underlying cause of the word length effect, it is considered to be one of the characteristic features of phonological STM and therefore it might be expected to remain intact in SD if the integrity of the phonological system is unaffected by this condition.

4.5.1 Method

Five patients and twelve control participants matched for age and educational level were included in this experiment. BS was not tested due to time constraints. Forty three-syllable words containing between six and eleven phonemes (mean = 7.5) and forty monosyllabic words containing between two and four phonemes (mean = 3.1) were selected (see Appendix 8). The two sets of words were matched closely on an item-by-item basis for word frequency (using data from Kucera & Francis, 1967; mean frequency long words = 91.8, short words = 91.0) and were assembled into lists of four items, yoked so that frequency-matched words appeared in the same positions in the lists for long and short words. Twenty lists of each word type were presented in a blocked fashion using an ABBA design. The words were read aloud at a rate of one word per second for immediate serial recall.

4.5.2 Results

4.5.2.1 Recall accuracy

Table 4.5 shows the number of items and lists recalled correctly for long and short words. Item recall was analysed using an ANOVA incorporating a within-subjects factor (word length) and a between-subjects factor (group: patients vs. controls). There were significant main effects of word length ($F(1, 15) = 24.04, p < 0.001$) and group ($F(1, 15) = 42.55, p < 0.0001$). There was no evidence of an interaction between these factors ($F(1, 15) < 1$), suggesting that the effect of word length did not differ greatly for the patients and controls. The patients' recall was generally below the normal range for both long and short words because they made numerous phonological errors on both sets of items. EK recalled significantly more short than long items ($t(38) = 3.29, p < 0.01$), as did KI ($t(37) = 2.65, p < 0.05$). SJ, JT and GT did not show any significant effects of word length (SJ: $t(38) < 1$; JT: $t(37) < 1.36$, n.s.; GT: $t(37) < 1$). As in the previous experiment, however, the effect of word length was very variable in the controls (consistent with the findings of Logie, Della Sala, Laiacona, Chambers & Wynn, 1996) and did not always reach significance ($t(25\text{-}38) = 5.86$, p < 0.0001 to $t < 1$). The size of the word length effect was within the control range for every patient.

Table 4.5: *Recall of long and short words*

		SJ	EK	KI	JT	GT	Control mean (range)
	Short	63.8*	77.5*	77.5*	66.3*	57.5*	91.3 (83 – 99)
Items	Long	56.3	51.3*	58.8	55.0	53.8*	77.7 (55 – 89)
	Difference	7.5	26.3	18.8	11.3	3.8	13.5 (3 – 34)
	Short	25*	45	35	30	5*	70.4 (30 – 95)
Lists	Long	15	10	10	5	15	41.7 (5 – 65)
	Difference	10	35	25	25	-10	28.8 (-20 – 60)

* denotes abnormal performance. Figures show percentage of items correct.

4.5.2.2 Recall errors

The influence of word length on the patients' errors was examined to cast further light on the processes underlying the word length effect. Incorrect responses were classified as 'phonological' if they contained at least half of the phonemes present in a target word. Other errors, e.g., omissions, order errors, repetitions and intrusions of items from previous lists, were classified as 'non-phonological'. Table 4.6 shows the number of phonological and non-phonological errors for patients and controls on short and long words. The patients made roughly normal numbers of non-phonological errors but inflated numbers of phonological errors that fell outside the control range, for both long and short words. Increases in word length were predominantly associated with increases in non-phonological but not phonological errors, for both patients and controls. Consequently, the balance of phonological to non-phonological errors varied across short and long words for SJ ($\chi^2(1) = 9.61, p < 0.01$), EK ($\chi^2(1) = 4.32, p < 0.05$) and JT ($\chi^2(1) = 11.01, p < 0.001$). This error difference between short and long words did not reach significance for GT ($\chi^2(1) = 2.48$, n.s.) or KI ($\chi^2(1) < 1$). The finding that word length predominately affected the occurrence of non-phonological rather than phonological errors suggests that the patients' difficulties in maintaining the coherence of items in STM did not interact with the amount of phonological material that was presented to them. This result is reminiscent of the finding that list length did not interact with the degree of phonological disintegration in ISR, obtained in Chapter 2.

Table 4.6: *Errors on long and short words*

		SJ	EK	KI	JT	GT	Control mean (range)
Short	Phonological	25*	11*	15*	26*	22*	2.9 (0 – 8)
	Non-phonological	4	7	3	1	12*	4.1 (0 – 8)
Long	Phonological	16*	11*	23*	20*	16*	3.1 (1 – 5)
	Non-phonological	19	28	10	16	21	14.8 (6 – 33)

* denotes abnormal performance. Figures indicate number of errors.

4.5.3 Discussion

Most of the patients exhibited effects of phonological similarity and word length in ISR that were within the normal range, consistent with them having normal phonological coding and rehearsal processes in verbal STM. It remains possible that phonological similarity and word length effects are slightly reduced in size in SD patients on average, although there was no conclusive evidence of this in the performance of individual patients. The word length effect in the patients appeared to result from a normal increase in the number of non-phonological errors on the longer words. Interestingly, word length did not affect the number of phonological errors committed by the patients. Therefore, it seems that although the patients had difficulty maintaining the phonological integrity of the words in STM, the amount of phonological material to be recalled did not increase the likelihood of phonological breakdown (see Chapter 2 for a similar conclusion).

4.6 Repetition of single multisyllabic nonwords

The patients and ten controls matched for age and educational level were tested on the Children's Test of Nonword Repetition (CN Rep; Gathercole et al., 1994), in order to investigate the suggestion that nonword repetition is impaired in SD. This test involves the immediate repetition of 40 nonwords, ranging in length from two to five syllables. Table 4.7 shows the scores obtained by the patients and controls. The five mildest patients, SJ, BS, EK, KI and JT were able to repeat the majority of items correctly, whereas GT showed some weakness on this test. It is important to note, however, that the performance of the controls was highly variable. Most of the controls made very few errors but several controls performed surprisingly poorly, possibly because they had some degree of mild hearing loss. The patients were within the normal range with the exception of GT, whose performance fell slightly below the poorest control score. GT showed little effect of item length and made an unusually large number of errors on the short items, consistent with the notion that his hearing loss was responsible for many of his errors.

Table 4.7: *Scores on the CN Rep test (Gathercole et al., 1994)*

	Max	SJ	BS	EK	KI	JT	GT	Control median score (range)
Total	40	35	36	38	32	28	21*	35 (23 – 39)
Short items (2-3 syllables)	20	18	19	20	18	19	9*	19 (14 – 20)
Long items (4-5 syllables)	20	17	17	18	14	9	12	17 (8 – 19)

* denotes abnormal performance. Figures indicate number of items correct.

In summary, the results of the CN Rep test suggest that nonword repetition was largely intact in these SD patients although the patient with the severest semantic deficit was mildly impaired. Interestingly, the performance of Knott et al.'s (1997) patient, AB, did not fall below the range of scores obtained for the control participants, although his performance may have been impaired in comparison with his pre-morbid abilities. The following experiment, which required lists of monosyllabic nonwords to be recalled in order, examined STM for nonwords in more detail.

4.7 Serial recall of monosyllabic nonword lists

4.7.1 Method

The six patients and ten control participants matched for age and educational level were asked to recall lists of monosyllabic CVC nonwords and words (taken from Gathercole, Pickering, Hall, & Peaker, 2001). The words were not selected according to the patients' knowledge of them. The patients were tested on lists containing two to four items and the controls were additionally tested on five-item lists. The words and nonwords, which were not repeated in the course of the experiment, were blocked using an ABBA design and were presented auditorily at a rate of one item per second for immediate serial recall.

4.7.2 Results

4.7.2.1 Recall accuracy

Table 4.8 shows the number of words and nonwords recalled by the patients and controls at each list length. The data were analysed using an ANOVA incorporating two within-subjects factors (list length and lexicality) and one between-subjects factor (group: patients vs. controls). The main effect of group approached significance ($F(1, 14) = 3.92$, $p = 0.07$). There were significant main effects of both lexicality ($F(1, 14) = 138.75$, $p < 0.0001$) and list length ($F(2, 28) = 97.54$, $p < 0.0001$), indicating that recall was better for the words than the nonwords and that percentage recall declined as list length increased. The interaction between lexicality and group reached significance ($F(1, 14) = 8.96$, $p < 0.01$). The patients' recall was impaired for words but not nonwords: the controls recalled the words more accurately than the patients, but the recall of nonwords did not differ across the groups (planned comparisons; $t(8) = 4.65$, $p < 0.01$ and $t(11) < 1$ respectively). In addition, there were significant interactions between length and group ($F(2, 28) = 3.77$, $p < 0.05$), lexicality and length ($F(2, 28) = 6.45$, $p < 0.01$) and lexicality, length and group ($F(2, 28) = 3.59$, $p < 0.05$). These interactions appeared to be caused by ceiling effects: the controls' recall of words but not nonwords was at ceiling on the shortest list length. In line with this suggestion, Bonferroni t tests showed that word recall was significantly better for the controls than the patients on lists containing three items ($t(14)$

= 4.08, $p < 0.01$) and four items ($t(8) = 5.35$, $p < 0.01$) but not on lists containing two items ($t(6) = 1.09$, n.s.). Nonword recall did not differ for the patients and controls at any list length ($t(9\text{-}14) < 1$).

For every patient, word recall was below the normal range on some list lengths, whereas only GT's nonword recall fell below the normal range. Furthermore, all the patients showed some reduction in the size of the lexicality effect. For the controls, the recall difference between words and nonwords on four item lists was between 60% and 28% (mean = 42%). For the patients, the size of this difference was between 10% and 23% (mean = 16%). A significant advantage for words over nonwords occurred for KI ($t(76) = 2.00$, $p < 0.05$) and GT ($t(70) = 2.41$, $p < 0.05$). The lexicality effect approached significance for SJ ($t(76) = 1.88$, $p = 0.06$) and did not reach significance for BS ($t(77) < 1$), EK ($t(78) < 1$) or JT ($t(74) = 1.57$, n.s.), combining the data from different list lengths. Without exception, the control participants showed very substantial effects of lexicality ($t(51\text{-}72) = 4.12 - 5.96$, $p < 0.001$).

Table 4.8: S*ingle-syllable words and nonwords recalled by patients and controls*

List length		SJ	BS	EK	KI	JT	GT	Control mean (range)
2	Words	100	100	100	85*	95	80*	98 (90 – 100)
	Nonwords	90	95	100	65	75	45*	82 (50 – 100)
3	Words	83	83	83	80	73*	73*	93 (77 – 100)
	Nonwords	43	67	77	53	47	47	57 (27 – 83)
4	Words	43*	63*	65*	48*	40*	45*	81 (70 – 98)
	Nonwords	23	53	43	28	30	33	40 (23 – 63)
5	Words	-	-	-	-	-	-	60 (42 – 78)
	Nonwords	-	-	-	-	-	-	20 (8 – 36)

* denotes abnormal performance. Figures indicate percentage of items recalled.

4.7.2.2 Recall errors

Table 4.9 shows the types of errors that occurred for patients and controls on words and nonwords, combining across list lengths that all participants were tested on. Errors were categorised as phonological if they contained at least half of the phonemes present in a target word. Omission errors occurred when fewer items were recalled than were presented. Order errors, repetitions of target items and intrusions of items from previously presented lists occurred infrequently and were placed in a single category of 'other' errors. Responses were placed in this category even if they met the criterion for a phonological error. Incorrect responses that could not be categorised as order errors, repetitions, intrusions or phonological errors were classified as 'unrelated'. In reality, these errors generally did preserve some of the phonemes of the target items, although fewer than 50%.

Table 4.9: *Errors on single-syllable words and nonwords*

		SJ	BS	EK	KI	JT	GT	Control mean (range)	
	List length	2-4	2-4	2-4	2-4	2-4	2-4	2-4	5
Words	Phonological	22*	13*	13*	27*	26*	28*	6.2 (1 – 11)	6.5 (3 –10)
	Unrelated	6*	1	4	3	4	5*	1.5 (0 – 4)	2.5 (0 – 9)
	Omission	0	6	1	0	0	0	2.1 (0 – 6)	9.0 (2 – 18)
	Other	0	0	1	0	3	1	0.3 (0 – 1)	2.0 (0 – 6)
Nonwords	Phonological	38	24	24	43*	36	33	29. 0 (29 – 40)	22.0 (16 – 27)
	Unrelated	12	1	6	7	13	21	7.9 (0 – 24)	9.4 (1 – 19)
	Omission	0	5	0	0	0	0	2.7 (0 – 15)	7.9 (2 – 18)
	Other	0	0	0	0	0	0	0.8 (0 – 7)	0.2 (0 – 1)

* denotes abnormal performance on lists containing 2 to 4 items. Figures show total number of errors for each list length.

All of the patients made frequent phonological errors in word recall that fell outside the normal range, consistent with the view that semantics makes an important contribution to the coherence of items in STM. In contrast, the patients made far fewer errors in the other categories, and the frequency of these errors largely fell within the normal range. In nonword recall, the number of phonological errors was substantially larger for the controls and consequently, with the exception of KI, the patients' errors in this category did not exceed the normal range.

An ANOVA was used to examine the numbers of phonological and non-phonological errors (including unrelated errors, omissions and other errors) made by the patients and controls on words and nonwords. This analysis collapsed across list length: as the controls but not the patients were tested on lists of five items, errors were expressed as a proportion of the number of items presented. The three-way interaction between participant group, lexicality and error type was significant ($F(1, 14) = 5.28$, $p < 0.05$). The controls showed a different balance of phonological to non-phonological errors for words and nonwords ($F(1, 9) = 32.60$, $p < 0.001$), whereas the patients did not ($F(1, 5) = 3.41$, n.s.). The controls made more phonological than non-phonological errors in nonword recall ($t(9) = 4.41$, $p < 0.01$) but did not show this preponderance of phonological errors in word recall ($t(9) = -1.90$, n.s.). The patients, in contrast, made a greater number of phonological than non-phonological errors in their recall of both words ($t(5) = 5.05$, $p < 0.01$) and nonwords ($t(5) = 6.54$, $p < 0.01$). In line with this pattern, the interaction between error type and lexicality reached significance ($F(1, 14) = 24.86$, $p < 0.0001$: there were more phonological errors in nonword than word recall). In addition, there was a significant interaction between participant group and error type ($F(1, 14) = 14.67$, $p < 0.01$: the patients made more phonological errors than the controls).

In the word recall task, the controls showed a strong bias to produce real word responses when they made phonological errors (combining across participants, 112/127 errors were real words). The patients showed this bias to a lesser extent (82/129 errors were real words), and as a result, the balance of word to nonword responses in the word recall task was different for the patients and controls ($\chi^2(1) = 19.82$, $p < 0.0001$). In the nonword recall task, the controls no longer showed a strong bias to produce real word responses (258/516 errors were real words) and the patients made a similar mix of word and nonword responses (88/198 errors were real words). Consequently, the balance of word to nonword responses did not differ between the patients and controls for nonword recall ($\chi^2(1) = 1.55$, n.s.).

4.7.3 Discussion

The patients' repetition of single multisyllabic nonwords taken from the CN Rep test (Gathercole et al., 1994) was largely intact, as was their recall of strings of monosyllabic nonwords. In contrast, the patients' recall of real words was substantially impaired and was characterised by an abnormally large number of phonological errors, similar to those observed in previous studies (McCarthy & Warrington, 1987; Patterson et al., 1994). Although these words were not selected according to the patients' understanding of them, it is likely that their comprehension was at least partly compromised, and therefore the phonological errors may have arisen because the semantic system was unable to constrain the phonological representations in STM in the normal way. Both the patients and the controls made a substantial number of phonological errors in their recall of nonwords, which by definition largely lack semantic support. The number of phonological errors made by the patients in nonword recall was for the most part within the normal range. However, there was some suggestion that the patients with the greatest semantic deficits may have been mildly impaired at both repeating single multisyllabic nonwords and lists of monosyllabic nonwords. It remains possible, therefore, that the cortical atrophy underlying SD impinges on the phonological system to some extent. Alternatively, the patients' nonword recall difficulties could have resulted from a reduction in the usual lexical-semantic support for nonwords in verbal STM. The following experiments aimed to examine the extent to which the patients' semantic impairments impinged on their nonword recall.

4.8 Recall of nonwords phonologically similar to semantically known and degraded words

Several studies have demonstrated better recall of words that individual SD patients still understand relatively well, compared with words that are more semantically degraded (e.g., Knott et al., 1997, 2000; Patterson et al., 1994). In this study, nonwords were constructed that were phonologically similar to known and degraded words in order to investigate the impact of semantic knowledge on nonword recall. Nonwords derived from known words should be recalled more accurately than nonwords derived from degraded words if stable semantic knowledge of phonologically similar words makes a contribution to nonword recall. The semantic binding hypothesis of Patterson and colleagues (1994) allows for such a contribution, as the bi-directional connections between phonological and semantic nodes might produce a stabilising influence on phonological representations even for nonwords. It follows that the stabilising semantic-phonological interaction will be graded by the degree of conceptual dissolution in the SD cases. In line with this prediction, one previous study found fewer phonological errors for nonwords derived from known words compared with nonwords derived from degraded words in a patient with herpes simplex encephalitis, although there was no difference between the two sets of nonwords in recall accuracy (Caza, Belleville, & Gilbert, 2002).

4.8.1 Method

Five SD patients and ten controls matched for age and educational level participated in this experiment, allowing two controls to be tested on the material presented to each patient. BS was not included due to limitations on testing time. Sets of semantically known and degraded words were selected for each patient using naming, definition and synonym judgement tests. The patients were asked to name 100 pictures from the Snodgrass set, as well as thirteen colours and twenty body parts, and to provide definitions for the same items. Naming attempts were considered to be correct when the patients produced the appropriate label for a picture. Definitions were considered to be correct when they contained enough specific information to allow the item to be identified from its description. Items that were both named and defined correctly were classified as known, and items that were neither named nor defined correctly were classified as degraded. The patients were also tested on a multiple choice synonym judgement test in which they were asked questions like "which word is closest in meaning to rogue: scoundrel, polka or gasket?" The test was administered twice on two separate occasions. Known items were responded to correctly on both occasions, and degraded items were responded to incorrectly on both occasions. The known and degraded words were matched for syllable length and word frequency as closely as possible on an item-by-item basis using data from Celex (Baayen, Piepenbrock, & van Rijn, 1993) and Kucera and Francis (1967). While this methodology should ensure a semantic difference between words assigned to the known and degraded categories, semantic degradation is thought to be a continuous variable. Consequently, the known words may not have been entirely intact and the degraded words may not have been entirely forgotten. Appendix 9 gives mean word frequency, length, imageability ratings and set size for each patient's known and degraded words.

Nonwords were constructed from the known and degraded words in two different ways. First, the onsets were switched between the items in each set to form monosyllabic and multisyllabic nonwords. For example, the words 'kangaroo' and 'strawberry' were used to produce the nonwords 'strangaroo' and 'kawberry'. Occasionally, it was necessary to create nonwords by replacing rather than switching onsets when the onsets of all the other words in the set produced real words and not nonwords. A second set of nonwords was constructed by exchanging the initial syllables of multisyllabic words. For example, the words 'pineapple' and 'strawberry' were combined to form the nonword 'pineberry', and the words 'motorbike' and 'screwdriver' were combined to form the nonword 'moe-driver'. This second method could not be used with single syllable words. Consequently it was not possible to test JT, as very few multisyllabic items were available for him. It was also not possible to test SJ in this part of the experiment, due to limitations on testing time.

Lists of known and degraded words and nonwords were assembled by selecting monosyllabic items at random without replacement until all the items had been used, and then repeating this process with the two-syllable and three-syllable items. Each list was therefore generally composed of items of a particular syllable length, although some lists contained a mixture of one and two syllable items and two and three syllable items. For convenience, these lists were grouped in the analysis with the pure lists that they were most similar to. This list construction process was repeated three times, so that over the course of the test, most items were presented in three separate lists. As the size of the word sets varied between the patients, each patient was tested on a different number of lists. The known and degraded lists were presented alternately, and the lists were yoked so that an item in a particular position in a known list and the corresponding item in the following degraded list formed a frequency-matched pair. The words and nonwords were tested in separate blocks, with the nonwords presented first. The word and nonword items were also yoked so that nonwords appeared in the same list positions as the words they were derived from. The patients were tested on lists containing three and four items. EK was also tested on lists of five words. The controls were additionally tested on five item lists and lists of six words. As the word sets were not always divisible by three, five and six, it was necessary to repeat a small number of items twice or four times. The items were read aloud at a rate of one word per second for immediate serial recall.

4.8.2 Results

4.8.2.1 Recall accuracy

Table 4.10 shows the percentage of words and nonwords recalled correctly by the patients and controls, averaging across lists containing three and four items, on which every participant was tested. Every patient's recall of the degraded words fell below the control range. SJ, KI and GT also showed some impairment of known word recall, possibly because their knowledge of these words was not entirely intact. In contrast, EK and JT recalled the known words at a normal level. All five patients recalled the nonwords derived from known words at a relatively normal level, whether they were constructed by exchanging onsets or initial syllables. EK and GT's recall of both sets of degraded nonwords was below the normal range, however, and SJ, KI and JT's recall of nonwords derived from degraded words was at the bottom of the normal range.

Table 4.10: *Percentage of known and degraded words and nonwords recalled by patients and controls*

		SJ	EK	KI	JT	GT	Control mean (range)
Words	Known	78.0*	87.5	85.6*	92.1	86.1*	94.7 (88 – 98)
	Degraded	64.0*	73.1*	72.7*	82.5*	64.2*	94.1 (85 – 100)
Nonwords: onsets exchanged	Known	46.8	40.7	53.7	57.0	37.6	53.7 (28 – 77)
	Degraded	40.9	32.4*	40.3	39.5	17.0*	51.6 (35 – 68)
Nonwords: syllables exchanged	Known	NT	36.8	51.4	NT	33.3	35.3 (22 – 56)
	Degraded	NT	19.4*	50.7	NT	23.7*	39.2 (28 – 58)

Note: Table combines data from three and four item lists.

* denotes abnormal performance

Figure 4.1 shows the number of known and degraded words recalled by the patients and controls collapsing across the full range of list lengths. The known words were recalled more accurately than the degraded words by SJ ($t(99) = 2.45$, $p < 0.05$), EK ($t(155) = 3.21$, $p < 0.01$), KI ($t(107) = 2.84$, $p < 0.01$) and GT ($t(76) = 4.22$, $p < 0.0001$) but not by JT ($t(45) = 1.64$, n.s.), when the data from every list length was combined. However, JT's recall was at ceiling for both the known and the degraded words on three-item lists and he showed a significant known-degraded difference when four-item lists were analysed separately ($t(19) = 2.41$, $p < 0.05$). None of the controls showed a significant known-degraded recall difference, both when the full range of list lengths were included in the analysis ($t(96\text{-}202) < 1.61$, n.s.) and when the analysis was restricted to five- and six-item lists ($t(41\text{-}76) < 1$).

Figure 4.1: *Recall of known and degraded words and nonwords of different syllable lengths*

Fig. 4.1a: *Words*

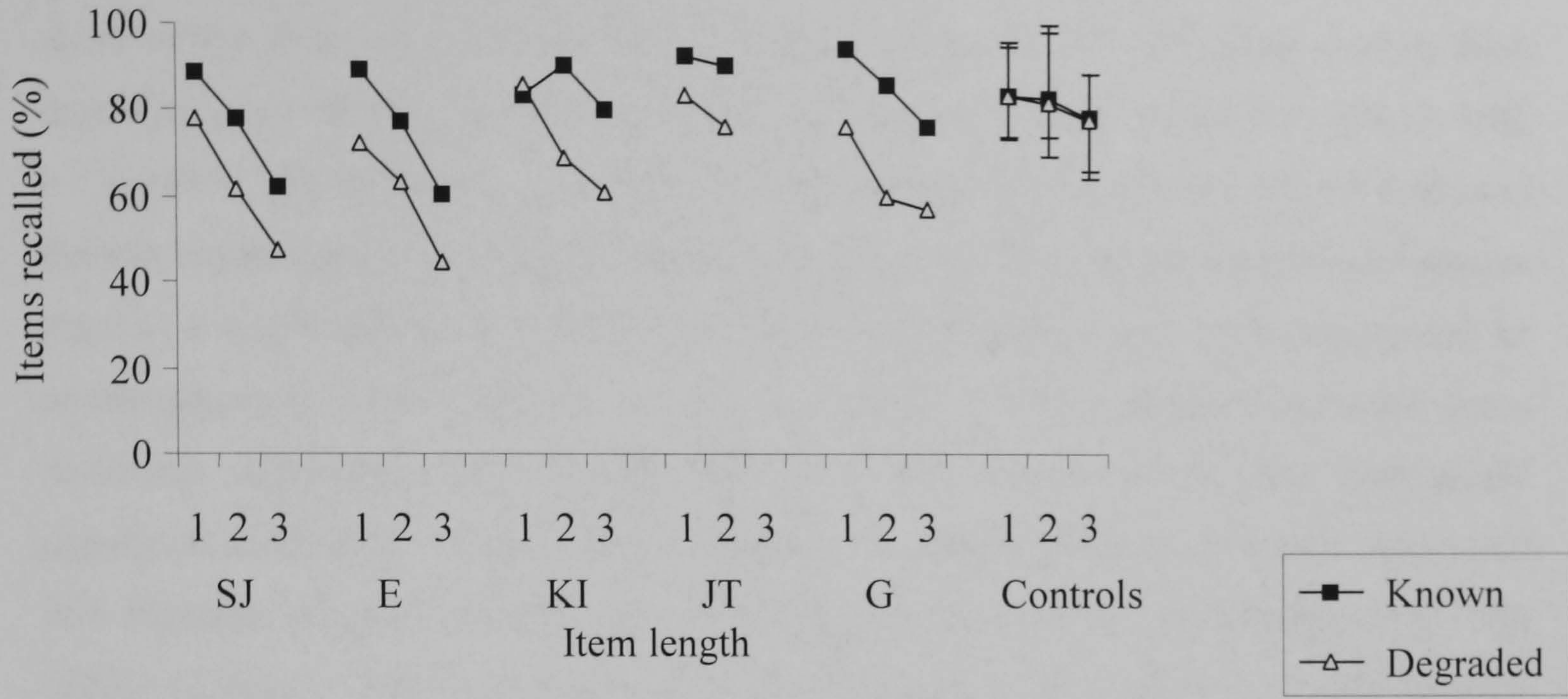
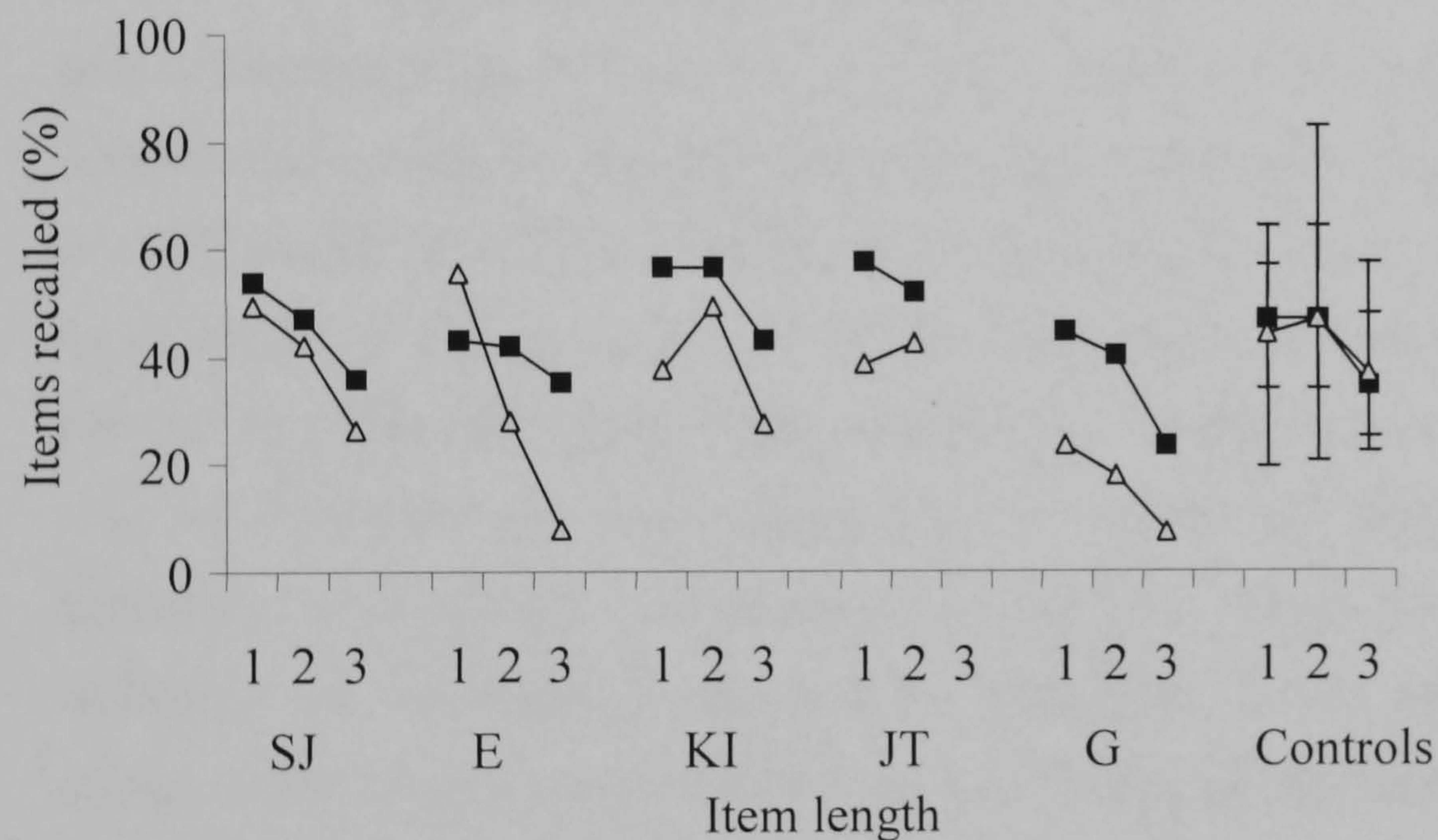
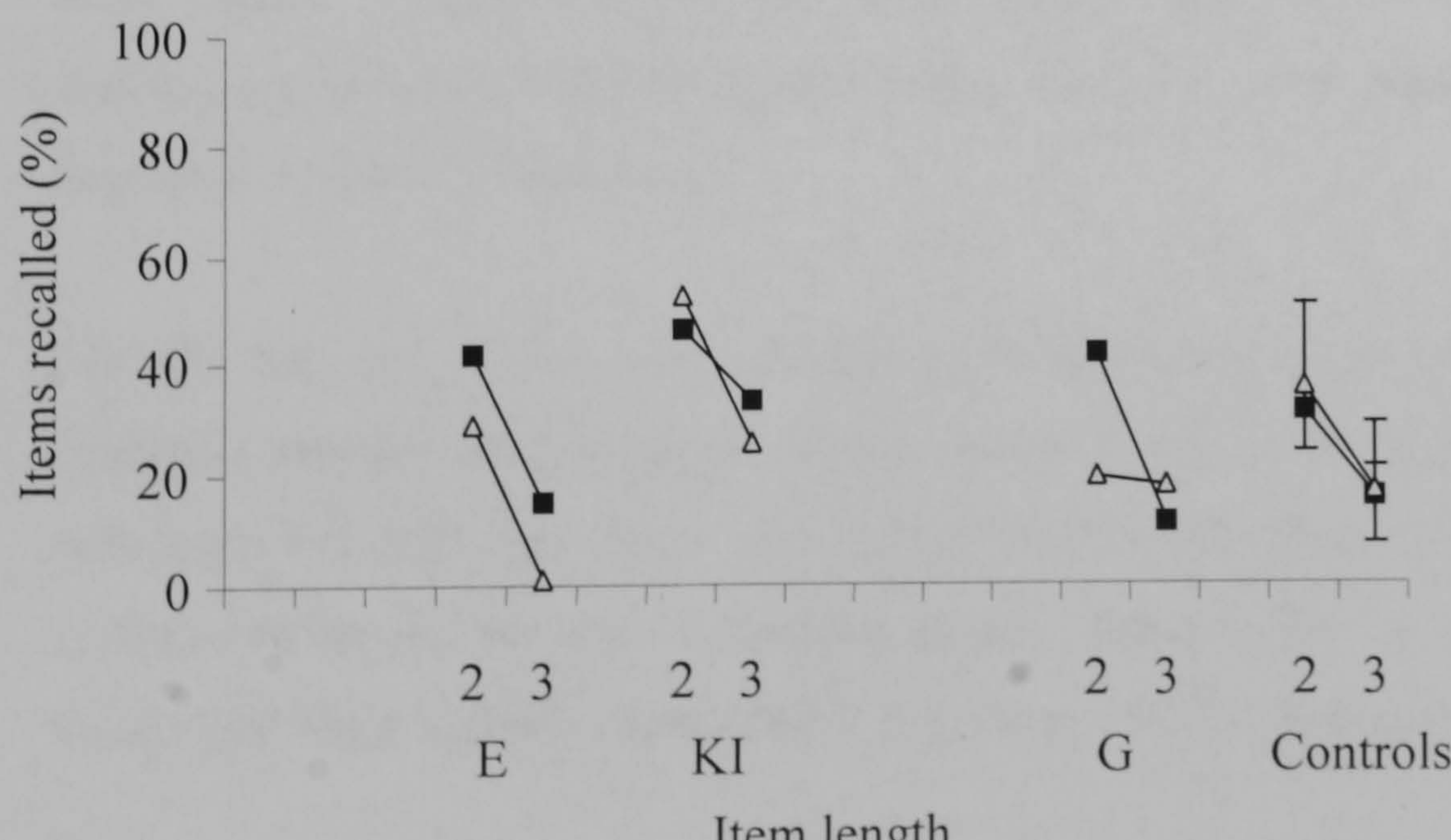


Fig. 4.1b: *Nonwords constructed by switching onsets between*



Note:
Error bars show control range. Data are combined across all list lengths tested, including additional longer lists for controls.

Fig. 4.1c: *Nonwords constructed by switching initial syllables between*



Turning to the nonwords constructed by exchanging onsets, KI, JT and GT showed better recall of the items derived from known words, compared with the items derived from degraded words (KI: $t(120) = 2.64, p < 0.01$; JT: $t(63) = 2.33, p < 0.05$; GT: $t(86) = 4.28$, $p < 0.0001$). This difference approached significance for EK ($t(120) = 1.74, p = 0.08$) and did not reach significance for SJ ($t(106) = 1.03$, n.s.). None of the control participants showed a recall difference between the known and degraded nonwords constructed by exchanging onsets (all $t(88\text{-}166) < 1.33$, n.s., except for one participant who showed a numerical advantage for the degraded over the known words that approached significance, $t(160) = 1.76, p = 0.08$). Similarly, the recall difference between the known and degraded items constructed by exchanging initial syllables was significant for EK ($t(80) = 2.91, p < 0.01$) but did not reach significance for KI ($t(82) < 1$) or GT ($t(64) = 1.74, p = 0.09$). As before, the control participants recalled the known and degraded items at an equivalent level (all $t(83\text{-}112) < 1.72$, n.s., except for two participants who showed a numerical advantage for the degraded words that approached significance, $t(108) = 1.79$, $p = 0.08$ and $t(111) = 1.92, p = 0.06$). In summary, all five patients consistently showed a recall advantage for known words over degraded words, and some patients (EK, KI, JT and GT) also showed superior recall of nonwords derived from known words compared with nonwords derived from degraded words. KI and GT showed a known-degraded difference for nonwords constructed by exchanging onsets, but did not show such a difference for nonwords constructed by exchanging initial syllables. The nonwords constructed by exchanging syllables were less phonologically related to individual known and degraded words, possibly reducing the size of the known-degraded difference for these items. In contrast, EK did show better recall of nonwords constructed by exchanging syllables between known words, compared with nonwords constructed from degraded words in a similar way.

For EK, the size of the recall difference between nonwords derived from known and degraded words varied with the length of the items to be recalled (see Figure 4.1b). Although EK did not show an overall known-degraded difference for nonwords constructed by exchanging onsets (see above), she did show a significant difference on the longest three-syllable items ($t(23) = 2.94, p < 0.01$). The known-degraded difference approached significance for two syllable items ($t(52) = 1.90$, $p = 0.06$) but was not significant for one-syllable items ($t(37) = 1.27$, n.s.). The other patients' recall of nonwords constructed from known and degraded words by exchanging onsets was less consistently affected by item length. KI, JT and GT all showed a significant known-degraded difference for one-syllable nonwords (KI: $t(39) = 2.07$, $p < 0.05$; JT: $t(52) = 2.43$, $p < 0.05$; GT: ($t(27) = 2.32$, $p < 0.05$). GT also showed a significant known-degraded difference for two-syllable items ($t(35) = 3.20$, $p < 0.01$) and three-syllable items ($t(18) = 2.08$, $p = 0.05$). KI did not show a significant known-degraded difference for two-syllable items ($t(52) < 1$) or three-syllable items ($t(19) = 1.78$, $p = 0.09$). JT also failed to show a known-degraded difference for two- and three-syllable nonwords ($t(8) < 1$), although few of these longer items were available for testing. SJ was the only patient who failed to show a significant known-degraded recall difference for nonwords constructed by exchanging onsets at every word length (one-syllable items: $t(34) < 1$; two-syllable items: $t(46) < 1$; three-syllable items: $t(21) = 1.04$, n.s.).

4.8.2.2 Recall errors

Tables 4.11, 4.12 and 4.13 show the types of errors that were made by the patients and controls in the recall of real words (Table 4.11), nonwords constructed by exchanging onsets (Table 4.12) and nonwords constructed by exchanging syllables (Table 4.13). These data combine the errors made on every list length, including the longer list lengths that the controls but not the patients were tested on, as the controls made very few word errors on the shorter list lengths. As in the previous experiment, errors were classified as phonological errors, omissions, 'other' errors (order errors, intrusions, repetitions), and 'unrelated' errors (if they could not be placed in the previous three categories). Some of the phonological errors, in which participants recalled nonwords as the words they had been derived from, were additionally classified as 'source' errors.

Table 4.11: *Errors on known and degraded words*

	Error type	SJ	EK	KI	JT	GT	Control mean (range)
Known	Phonological	.67*	.41*	.74*	.78*	.78*	.17 (.05 – .36)
	Unrelated	.17*	.08*	.06	.11*	.04	.02 (0 – .06)
	Omission	.07	.38	.06	.00	.13	.49 (.28 – .76)
	Other	.10	.14	.13	.11	.04	.31 (.19 – .44)
Degraded	Phonological	.75*	.49*	.86*	.55*	.93*	.15 (.03 – .28)
	Unrelated	.10*	.12*	.02	.30*	.02	.02 (0 – .05)
	Omission	.07	.24	.02	.15	.05	.50 (.30 – .80)
	Other	.07	.15	.10	.00	.00	.33 (.13 – .51)

Note: Table combines data across every list length tested. Figures show the proportion of errors in each category

* denotes abnormal performance

Table 4.11 shows that the patients made an abnormally high proportion of phonological errors on both the known and degraded words. Phonological errors were the commonest type of word recall error for patients but were relatively uncommon among the controls. None of the patients showed a significant error difference between the known and degraded words, as phonological errors were relatively common for both ($\chi^2(3)$ from 1.34 to 5.12, n.s.). In contrast, the patients' errors were more normal for the nonwords constructed by exchanging onsets between known words (see Table 4.12). Both the patients and controls made frequent phonological errors on these nonwords, and consequently the patients' phonological errors did not exceed the normal range. The number of phonological errors on the nonwords constructed from degraded words did exceed the normal range for three out of five patients, consistent with the notion that a reduction in semantic support led to a reduction in phonological coherence for these nonwords in STM. However, none of the patients made significantly different types of errors on the known and degraded nonword items ($\chi^2(3)$ between 1.39 and 4.33, n.s.). For the nonwords constructed by moving syllables (see Table 4.13), the patients made a greater proportion of phonological errors compared with the controls, for both known and degraded items, and showed no significant differences between their errors on the known and degraded items ($\chi^2(3)$ between <1 and 4.18, n.s.). Source errors occurred relatively infrequently for both sets of nonwords and were largely equivalent for the patients and controls, although the patients occasionally exceeded the normal range.

Table 4.12: *Errors on known and degraded nonwords constructed by exchanging onsets*

	Error type	SJ	EK	KI	JT	GT	Control mean (range)
Known	Phonological	.79	.81	.87	.88	.76	.67 (.48 – .88)
	Source	.06	.03	.19*	.08	.07	.08 (.01 – .14)
	Unrelated	.18	.13	.10	.06	.21	.11 (.05 – .21)
	Omission	.03	.05	.02	.06	.03	.21 (.01 – .04)
	Other	.00	.00	.01	.00	.00	.01 (0 – .04)
Degraded	Phonological	.89*	.77	.91*	.93*	.80	.67 (.43 – .81)
	Source	.05	.03	.14	.09	.07	.09 (.04 – .14)
	Unrelated	.09	.18*	.09	.04	.15*	.10 (.06 – .14)
	Omission	.02	.05	.01	.03	.04	.21 (.05 – .44)
	Other	.00	.00	.00	.00	.00	.02 (0 – .04)

Note: Table combines data across every list length tested. Figures show the proportion of errors in each category.

* denotes abnormal performance

Table 4.13: *Errors on known and degraded nonwords constructed by exchanging initial syllables*

	Error type	EK	KI	GT	Control mean (range)
	Phonological	.76*	.80*	.83*	.59 (.44 – .68)
	Source	.11*	.04	.11*	.06 (.03 – .09)
Known	Unrelated	.15	.17	.14	.11 (.07 – .18)
	Omission	.09	.03	.03	.29 (.14 – .46)
	Other	.00	.00	.00	.01 (0 – .04)
	Phonological	.76*	.89*	.83*	.57 (.36 – .67)
	Source	.03	.08	.10*	.06 (.02 – .08)
Degraded	Unrelated	.13	.07	.15	.10 (.04 – .19)
	Omission	.11	.03	.02	.33 (.08 – .59)
	Other	.00	.01	.00	.01 (0 – .01)

Note: Table combines data across every list length tested. Figures show the proportion of errors in each category.

* denotes abnormal performance

The controls made substantially different types of errors on words and nonwords, principally because phonological errors were much more common for nonwords. Every control participant showed a highly significant error difference between words and nonwords constructed by exchanging onsets, for both known items ($\chi^2(3)$ between 19.96 and 143.35, $p < 0.0001$) and degraded items ($\chi^2(3)$ between 17.84 and 124.62, $p < 0.0001$). Similarly, every control participant showed a highly significant error difference between words and nonwords constructed by switching initial syllables, for both known items ($\chi^2(3)$ between 48.97 and 106.15, $p < 0.0001$) and degraded items ($\chi^2(3)$ between 34.73 and 99.02, $p < 0.0001$). The patients also typically showed significant errors differences between words and nonwords, although they tended to be rather weaker than those shown by the controls because substantial numbers of phonological errors occurred in both word and nonword recall. There was a significant error difference between known words and nonwords constructed by exchanging onsets for SJ ($\chi^2(3) = 9.91$, $p < 0.05$), EK

($\chi^2(3) = 58.00$, $p < 0.0001$), KI ($\chi^2(3) = 9.43$, $p < 0.05$) and GT ($\chi^2(3) = 10.13$, $p < 0.05$), but not JT ($\chi^2(3) = 4.81$, n.s.). Similarly, there was a significant error difference between degraded words and nonwords constructed by exchanging onsets for SJ ($\chi^2(3) = 12.17$, $p < 0.05$), EK ($\chi^2(3) = 49.30$, $p < 0.0001$), KI ($\chi^2(3) = 15.41$, $p < 0.05$), JT ($\chi^2(3) = 14.53$, $p < 0.001$) and GT ($\chi^2(3) = 8.91$, $p < 0.05$). In addition, all three patients tested showed a significant error difference between known words and nonwords constructed by exchanging syllables (EK: $\chi^2(3) = 39.59$, $p < 0.0001$; KI: $\chi^2(3) = 10.40$, $p < 0.01$; GT: $\chi^2(3) = 7.48$, $p < 0.05$) and two patients showed a significant error difference between degraded words and nonwords constructed by exchanging syllables (EK: $\chi^2(3) = 29.28$, $p < 0.0001$; KI: $\chi^2(3) = 6.50$, $p = 0.07$; GT: $\chi^2(3) = 8.28$, $p < 0.05$).

4.8.3 Discussion

All five patients were found to recall relatively well-known words better than semantically degraded words, consistent with the view that semantics makes an important contribution to the coherence of the phonological representations of words in STM. Furthermore, the ISR difference between known and degraded words remained substantial even for those patients who showed standard effects of phonological similarity and word length in ISR, intact nonword recall/repetition and normal phonological processing skills, suggesting that the finding of known-degraded recall differences is not dependent on additional phonological deficits (see Chapter 2 and Section 4.10). In addition, the patients recalled nonwords derived from relatively well-known words more accurately than nonwords derived from more semantically degraded words, suggesting that the semantic system may contribute to the phonological coherence of nonwords as well as words in STM. Although this effect was more marked for the patients with more severe semantic impairments, it was evident to some extent for every patient except SJ. There was no substantial effect of semantic status on the frequency of different error types, unlike in the study of Caza et al. (2002), although for three out of five patients, there was some suggestion that phonological errors occurred at an inflated rate for 'degraded' nonwords and at a more normal level for 'known' nonwords.

EK only showed a known-degraded difference in recall accuracy for longer multisyllabic nonwords. It seems likely that longer nonwords that are highly phonologically related to individual known words (and few other words) will produce some of the semantic activation associated with those real words, which will in turn feed back to the phonological system and support the coherence of longer nonwords in phonological STM. In contrast, shorter nonwords are likely to have several phonological neighbours that vary in their degree of semantic degradation, and consequently less clear differences may emerge. For example, if the word 'fork' was semantically degraded, recall of the nonword 'rork' might be supported by other phonologically similar words (e.g. 'walk', 'talk' etc) that were still relatively well understood. Having said this, several other patients (KI, JT and GT) did show significant known-degraded differences on the shorter nonwords, perhaps because their lexical-semantic deficits had markedly reduced the size of the nonwords' phonological neighbourhoods. These patients, who had particularly severe semantic impairments, may have essentially forgotten about a large number of words, making it less likely that nonwords derived from degraded words would be supported by other phonologically similar words.

Nonwords that were constructed by exchanging syllables between items may have produced more equivocal results than nonwords that were produced by exchanging onsets for similar reasons. The nonwords that were constructed by exchanging onsets remained more phonologically similar to the words they were derived from. The nonwords assembled by exchanging syllables may have shared their syllables with words that were both well understood and semantically degraded. Despite this, however, EK did show a significant known-degraded recall difference for these nonwords.

The results of this experiment suggest that semantic degradation can impact on the recall of nonwords. Therefore, it is not necessary to posit a separate phonological deficit to account for the impaired nonword recall observed in some SD patients. However, as SD patients are impaired at tasks like lexical decision, as well as at semantic tasks, it remains possible that a separate lexical impairment accounts for the poor nonword recall observed in SD, and this is likely to be the preferred interpretation of theorists who propose that separate lexical and semantic-level representations underpin language processing. In contrast, Patterson and colleagues assert that separate lexical-level representations are unwarranted (Patterson et al., 1996; Seidenberg & McClelland, 1989), leading to the prediction that the patients' difficulties in verbal STM and their markedly impaired comprehension result from the same central semantic impairment and not from dissociable impairments to independent conceptual and lexical representations. The final experiment described in this chapter attempted to address this issue.

4.9 Delayed copying of known and degraded items

If the verbal STM difficulties of SD patients result from a central semantic deficit and not from a separable lexical impairment, then it should be possible to demonstrate superior performance for the known over the degraded items using an entirely non-verbal task. Previous work has suggested that delayed copying is highly sensitive to the breakdown of semantic memory in SD (Bozeat et al., 2003). SD patients typically omit distinctive features from their delayed copies, resulting in more prototypical drawings. They also show a tendency to include incorrect features that are shared across a domain; for example, adding four legs to a duck because most animals have four legs. Therefore, this study examined delayed copying of pictures representing the known and degraded items used in the previous experiment, in order to determine whether a difference mirroring the known-degraded difference in ISR would emerge. If a single semantic deficit is viewed as underlying problems in both ISR and delayed copying, the same items should be impaired in the two tasks. If, in contrast, the ISR impairments of SD patients are underpinned by a lexical deficit that is independent of the semantic system, we should not see a correlation between the items impaired in ISR and delayed copying, even if both semantic and lexical representations are compromised in SD.

4.9.1 Method

SJ, EK, KI and JT participated in this study, which examined delayed copying of pictures from the Snodgrass set corresponding to the known and degraded items used in the previous experiment. It was not possible to test all of the known and degraded items, as some were not imageable or were not included in the Snodgrass set. Items were only included if a Snodgrass picture was available for both the known and degraded items in a matched pair. It was possible to test 20/32 item pairs for SJ, 15/36 for EK, 13/36 for KI and 11/20 for JT. The patients were allowed to study each picture for as long as they wished before it was hidden from view. The delay between viewing the picture and attempting to copy it was filled with a distracter task designed to disrupt visual-spatial working memory. This task involved making an immediate copy a geometric figure, similar to a simplified Rey figure, which was composed of a large rectangle, a square and two circles. On each trial, these shapes were arranged in a different configuration. The patients began making their delayed copies of the Snodgrass pictures as soon as the distracter task was completed. SJ, KI and JT copied a single distracter figure on each trial and EK copied two distracter figures, as pilot testing suggested her delayed copying showed little disruption with a single figure. The patients typically made no errors in their immediate copies of the distracter figures, in line with their intact performance on the Rey figure copy task (see Table 4.1). The known and degraded items in each pair were tested alternately.

4.9.2 Results

SJ, EK and JT produced a response on every trial, whether the picture being copied was from the known or the degraded set. KI, in contrast, produced a response on every known trial but failed to respond on 7/13 degraded trials. On these trials, he indicated that he had entirely forgotten the picture to be copied, including its overall shape.

The quality of each patient's known and degraded drawings was evaluated using a feature listing technique (Bozeat et al., 2003). For each target picture, a list of features was complied into a checklist. The presence or absence of particular features (e.g., body, head, limbs and tail) was recorded separately from modifiers of those features (e.g., body shape and size, long or short tail). The checklist was then used to identify features that had been correctly reproduced, omitted or incorrectly included by the patients. It should be acknowledged, however, that ideally, independent raters, blind to the source of the drawings should have been used to compare the patients' output with that of controls. The proportion of hits, misses and intrusions for each patient's known and degraded items are shown in Table 4.14, along with the total number of features identified. For all four patients, the proportion of hits to errors (misses and intrusions) was significantly higher for known items compared with degraded items (SJ: $\chi^2(3) = 18.54, p < 0.0001$; EK: $\chi^2(3) = 9.29, p < 0.01$; KI: $\chi^2(3) = 32.22, p < 0.0001$; JT: $\chi^2(3) = 7.19, p < 0.01$).

Table 4.14: *Delayed copying of known and degraded items*

	Error type	SJ	EK	KI	JT
	Hits	.84	.72	.47	.66
Known	Misses	.13	.25	.43	.28
	Intrusions	.03	.03	.10	.08
	Total features	256	193	142	150
	Hits	.69	.58	.19	.50
Degraded	Misses	.24	.34	.76	.39
	Intrusions	.07	.08	.06	.11
	Total features	335	220	220	135

Note: Hits, misses and intrusions are expressed as a proportion of the total number of hits plus errors.

4.9.3 Discussion

This experiment demonstrated that the ISR advantage for known over degraded items extended to a non-verbal delayed copying task. The same items were relatively well preserved and impaired in both ISR and delayed copying, suggesting that a central semantic deficit, which impinged on both the verbal and non-verbal domains, underpinned the patients' ISR impairments. Although it is conceivable that the atrophy underlying SD could encroach on independent semantic and lexical representations in such a way that lexical and semantic deficits would correlate across patients, this two-deficit account could not easily accommodate the finding that the same items are impaired in verbal and non-verbal tasks. This finding also suggests that the ISR difference between nonwords derived from known and degraded words resulted from the patients' semantic impairments, rather than from an independent lexical deficit, and therefore points to a genuine semantic contribution to nonword recall. One potential caveat should be noted, however. KI's frequent failures to produce a response on the degraded trials could have occurred because he struggled to retain the name of the object he was required to draw, and this difficulty may have been underpinned by a lexical deficit. Nevertheless, this explanation does not appear to provide a straightforward account of the pattern of omitted and intruded features that characterised SJ, EK and JT's copying of degraded items.

4.10 General Discussion

This chapter examined the performance of six patients with semantic dementia (SD) on a range of phonological processing and verbal STM tasks (summarised in Table 4.15), in order to explore the suggestion that semantics makes a major and necessary contribution to the coherence of items in verbal STM. According to this viewpoint, pure semantic impairments will result in phonological breakdown in immediate serial recall (ISR) tasks because of the normal interactive nature of the phonological and semantic systems, and it is not necessary to posit additional phonological or lexical impairments to account for the poor verbal STM performance of SD patients (Patterson et al., 1994). Alternatively, it has been argued that SD patients who show poor ISR performance have additional lexical-phonological deficits (McCarthy & Warrington, 2001).

In a range of tasks requiring phonological segmentation, discrimination between minimally different pairs of items, and rhyme judgement and production, the patients with the least impaired semantics performed normally, whereas the patients with more severe semantic deficits showed some mild weaknesses. Although this pattern is consistent with the view that the atrophy underlying SD ultimately impacts on phonological as well as semantic representations, an alternative possibility is that the patients' semantic impairments made it difficult for them to perform these tasks successfully. Given that they required the maintenance and manipulation of phonological representations, it seems likely that these tasks would not have been immune from the impact of semantic degradation if semantics does play a major role in constraining phonological activation. This may have been particularly true of the rhyme judgement task; more widespread deficits may have been observed on this task (see Table 4.14) because it required the phonology of two words to be maintained accurately while they were compared.

Table 4.15: *Summary of results for each patient*

	SJ	BS	EK	KI	JT	GT
Phoneme segmentation: addition	✓	✓	×	×	✓	×
Phoneme segmentation: subtraction	✓	✓	✓	✓	✓	×
Minimal pairs	✓	✓	✓	✓	✓	?
Rhyme judgement	×	✓	×	×	✓	NT
Rhyme production	✓	✓	✓	✓	×	NT
Phonological similarity effect in ISR	✓	✓	✓	?	×	?
Word length effect in ISR	✓	✓	✓	✓	✓	✓
CN Rep	✓	✓	✓	✓	✓	×
Known/degraded difference in word ISR	P	NT	P	P	P	P
Known/degraded difference in nonword ISR	A	NT	A[1]	P	P	P

Severity of SD increases across patients from left to right

✓ denotes intact performance, × denotes impaired performance, ? denotes marginal impairment, NT = not tested

P denotes presence of effect, A denotes absence of effect

[1] Effect on longer nonwords only

The effect of phonological similarity and word length on the patients' ISR performance was also examined and found to be within the normal range for all but the most severely impaired patients. These findings suggest that the patients with milder semantic impairments, in common with normal participants, relied heavily on a phonological code in ISR. Although the size of the phonological similarity effect may have been reduced in the more severely impaired patients, several recent studies have found that phonological similarity interacts with lexicality in ISR (Gathercole et al., 2001; Lian, Karlsen, & Winsvold, 2001) and consequently, semantic impairments might be expected to produce a reduction in the size of the phonological similarity effect in the absence of additional phonological deficits. In normal participants, phonological similarity effects are larger for words than for nonwords, suggesting that the phonological similarity effect may have its roots in the lexical-semantic binding process that operates for words and to a lesser extent for word-like nonwords. As this process is apparently disrupted in SD patients, the phonological similarity effect may be reduced in size.

Turning to the issue of ISR for nonwords, the majority of the SD patients showed normal recall of both single multisyllabic nonwords and strings of monosyllabic nonwords. Moreover, all the patients showed a reduction of the normal lexicality effect, suggesting that their semantic deficits had a larger impact on word than nonword recall. Again, there was some suggestion of a mild weakness in the patients with the most severe semantic impairments, which would be consistent with the disease process affecting phonological as well as semantic representations. Importantly, however, nonwords that were phonologically similar to relatively well known words were recalled more accurately than nonwords derived from semantically degraded words, particularly for the more impaired cases. This ISR difference, which points to a semantic contribution to nonword recall, could account for the nonword recall deficits of some SD patients.

These results are summarised in Table 4.15. The two mildest patients, SJ and BS, showed virtually normal performance in every task tapping phonological processing, whereas the other four cases showed some subtle difficulties in several of these tasks. These deficits were particularly marked for GT, who was the most severely semantically impaired case included in this study. Importantly, the presence or absence of a known-degraded difference in nonword recall mirrored the pattern of results obtained on the phonological tasks. The patients who showed mild deficits on the tasks designed to tap phonology also invariably displayed a known-degraded difference in ISR for nonwords. In contrast, SJ and EK, who had comparatively mild semantic impairments, did not consistently show better recall of nonwords that were derived from relatively well-understood words. This finding is consistent with the notion that the semantic deficits of the more severely impaired patients impacted on their ability to perform predominantly phonological tasks. In other words, it is hypothesised that mild semantic deficits have a clear effect on tasks that involve a considerable contribution from semantics as well as phonology (e.g., ISR of known and degraded words) but not on tasks that are largely underpinned by the phonological system (e.g., ISR of nonwords). In contrast, greater semantic deficits may have a measurable effect on predominately phonological tasks if the semantic and phonological systems are highly interactive. It should be noted, however, that given this interaction between phonology and semantics, mild phonological deficits might be expected to place a greater burden of processing on the semantic system, possibly accounting for the known-degraded difference in nonword recall observed for the more severely impaired patients.

The results presented here cannot rule out the possibility that phonological as well as semantic representations are compromised in the later stages of SD, although no conclusive evidence of such a deficit was obtained. The results strongly suggest, however, that the phonological coherence of semantically degraded words breaks down in SD patients even in the absence of additional phonological problems. All of the patients showed a significant ISR advantage for known over degraded words, even those patients who showed no signs of additional phonological impairment. Therefore, the results support the view that semantics makes a major contribution to the phonological stability of words in verbal STM, as suggested by Patterson et al. (1994). The findings are inconsistent with the claim that semantics makes little contribution to ISR performance except in the presence of phonological deficits. It seems likely that every patient showed a recall advantage for known words in this study, in contrast with the mixed findings of previous studies (Funnell, 1996; Lambon Ralph & Howard, 2000; McCarthy & Warrington, 1987, 2001), because set size was large for every patient. However, the incidence of phonological errors did not differ for the known and degraded words, in contrast with some previous studies (e.g., Knott et al., 1997; Patterson et al., 1994; see also Chapter 2), principally because the patients examined here made an abnormally large number of phonological errors on both the known and the degraded items. One possible explanation for this discrepancy is that items assigned to the 'known' category in this study were rather less well known than items assigned to that category in previous studies, given that continuous variation underlies the known-degraded distinction.

The ISR advantage for the known over the degraded items extended to a non-verbal delayed picture-copying task. The patients were able to reproduce more of the correct features when they made delayed copies of drawings that represented their known items, compared with their degraded items. As the same items were relatively well preserved in both ISR and delayed copying, it seems likely that a central semantic deficit, and not a separable lexical impairment, underpinned the patients' ISR impairments. A parallel argument was made in a recent study in which a very similar decline in performance was observed for lexical and object decision (Rogers, Lambon Ralph, Hodges, & Patterson, submitted) and these findings follow previous studies that strongly point to a central semantic deficit at the heart of this disorder (Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000; Lambon Ralph & Howard, 2000). This reasoning applies equally to ISR for words and nonwords; therefore, the patients' deficits in nonword recall apparently resulted from their marked semantic difficulties and not from any independent impairment of lexical representations. Consequently, the results of the current study point to a genuine semantic contribution to nonword recall.

The suggestion of a semantic contribution to nonword recall is a novel finding that requires some further discussion. There are at least two mechanisms by which this effect could occur. First, interactive models (e.g., N. Martin & Saffran, 1997; Patterson et al., 1994) appear to predict that nonwords closely resembling known words will be recalled more accurately than those resembling degraded words, as bi-directional connections between phonology and semantics should help to stabilise the nonword segments that overlap with known words. This semantic contribution to nonword recall is expected to be greatest for longer multisyllabic items, as long nonwords that are highly phonologically related to particular known words and few other words (e.g., 'strangeroo'

– a close neighbour of 'kangaroo') should produce strong and coherent activation of specific semantic representations relating to those individual words. EK did show a larger ISR difference between 'known' and 'degraded' nonwords for longer items and normal participants might also be expected to exhibit semantic effects for longer nonwords. In contrast, short nonwords are likely to have many phonological neighbours, all of which should produce activation in the semantic system. As a nonword like 'rork' will have several phonological neighbours with disparate meanings, e.g., walk, hawk, fork (i.e., the mappings between semantics and phonology are not systematic), it is less clear how semantic activation could helpfully constrain the phonological trace of these items.

Several of the more severely semantically impaired patients did show significant known-degraded differences on the shorter nonwords. The PDP framework (Patterson et al., 1994; Plaut & Kello, 1999) might allow for an effect of semantic impairment even for short nonwords as, according to this approach, stable phonological representations are acquired in the presence of semantics. For example, in the "model-T" framework of Plaut and Kello (1999), activation between acoustic, phonetic and semantic representations is accomplished through a common set of hidden units. In effect, these units and their associated connections to the surface representations form a "phonological" space. Given that there are no direct connections between acoustic and phonetic units, repetition of words and nonwords is achieved by passing activation through these hidden units. The nature and accuracy of repetition will reflect the 'topography' of the multi-dimensional space formed during the learning phase in the model. Given the high degree of systematicity between acoustic and phonetic representations, the representational space will predominantly reflect these associations (i.e., will form something like phonological representations). In addition, however, the model is required to transform these intermediate representations into meaning, and thus semantic memory will also influence the formation of this "phonological" space to at least some degree.

This framework provides a concrete instantiation of the semantic binding hypothesis (Patterson et al., 1994) in that phonological representations automatically interact with meaning. For words, ISR is likely to be more accurate when the corresponding semantic representations are relatively intact. In addition, once a patient's semantic impairment is sufficiently severe, the removal of semantic representations will have an impact, albeit subtle, on the generic "phonological" space formed across the connections to the hidden units. Given that this space is used for all items, both words and nonwords should be affected.

5 Lexical and semantic factors impact on phonological coherence in normal immediate serial recall

5.1 Introduction

The pattern of phoneme migration errors made by SD patients on semantically degraded words in ISR (see Chapters 2, 3 and 4) appears to offer an intriguing insight into the role of semantics in verbal STM. It appears that in healthy individuals, stable semantic representations help to maintain the phonological elements of words in the correct configuration. Patterson et al. (1994) reported that the recall errors of SD patients typically consisted of incorrect combinations of phonemes from different words that largely preserved onset/rime syllable structure (e.g., the onsets in the words 'mint, rug' were exchanged to produce the response 'rint, mug'). Treiman and Danis (1988) observed a similar pattern of phoneme migration errors in normal participants' recall of lists of nonwords (e.g., 'gir, vang, kus' recalled as 'gir, kang, vus'). In contrast, phoneme migrations are rarely observed in the ISR of healthy participants – instead, whole items are typically recalled in the wrong order (e.g., Henson, Norris, Page, & Baddeley, 1996; Pickering, Gathercole, & Peaker, 1998), particularly when a small pool of repeating items is used to construct the lists for recall (V. Coltheart, 1993; Gathercole, Pickering, Hall, & Peaker, 2001).

As noted in Chapter 1, this striking association between a lack of semantic support and the emergence of phoneme migration errors led Patterson et al. (1994) to propose that semantic constraints make a major contribution to the integrity of phonological representations in STM (see Section 1.3.1.2). According to this 'semantic binding hypothesis', the elements of words become associated in the phonological system because they are always activated together during speech production and comprehension.

As a result, they are more likely to emerge together in ISR. Similarly, because specific semantic activation frequently co-occurs with the phonology of words, semantic constraints encourage word phonemes to be produced in the correct order in ISR. According to this framework, there is no dedicated phonological STM system like the phonological loop that can operate independently of lexical and semantic knowledge. Instead, the language system, which incorporates lexical and semantic constraints, underpins phonological STM. This hypothesis posits distinct semantic and phonological representations but no separate lexical level, in line with the parallel distributed processing (PDP) models of, for example, Seidenberg and McClelland (1989). Nevertheless, both lexical and semantic knowledge are proposed to contribute to phonological stability. 'Lexical' constraints on phonology result from a combination of long-term learning of frequently co-occurring phonemes and stable associations between phonological and semantic representations.

This viewpoint differs somewhat from the dominant 'redintegration' account of the LTM contribution to verbal STM (Baddeley, Gathercole, & Papagno, 1998; Hulme, Maughan, & Brown, 1991; Hulme et al., 1997; Schweickert, 1993; see Section 1.3.3), although the two accounts make some similar predictions. The redintegration account proposes that two separate mechanisms underlie ISR performance. There is a rapidly decaying phonological STM store, which is initially inert to the effects of lexical and semantic factors, and a later reconstructive process that can only substantially improve the recall of words (although there may be influences of lexical neighbours on nonword recall: Roodenrys & Hinton, 2002). Redintegration compares the degraded STM trace of an item with separate long-term phonological-lexical representations, in order to reinstate the correct phonological activation during the process of recall. Although this reconstructive process is underpinned by phonological-lexical representations, the model can account for semantic effects in ISR by assuming that semantic activation contributes to the selection of lexical candidates for reconstruction (Poirier & Saint-Aubin, 1995). It is important to note that according to these theories, redintegration operates for individual items at the point at which they are recalled – therefore the redintegration of a particular item should not influence the recall of other list items.

It has been proposed that redintegration primarily benefits memory for item rather than order information (Gathercole et al., 2001; Poirier & Saint-Aubin, 1995, 1996; Saint-Aubin & Poirier, 2000). If the phonological trace of an item is degraded, redintegration should increase the probability of recalling the whole item and its constituent phonemes correctly but should not increase the probability of recalling the item in its correct serial position. In line with this prediction, several studies have found that lexical and semantic factors affect item identity but not item order errors (Gathercole et al., 2001; Hulme et al., 1997; Poirier & Saint-Aubin, 1995, 1996; Saint-Aubin & Poirier, 1999, 2000; Walker & Hulme, 1999). The semantic binding account makes less explicit predictions about the effect of lexical and semantic variables on item order errors, but clearly does predict that these variables should affect the number of phoneme order errors. There is little relevant data on this point, although Gathercole et al. (2001) found that in nine-year old children, lexicality primarily affected the occurrence of item rather than order errors, both at the level of whole items and individual phonemes. This finding appears to be at odds with the frequent phoneme migration errors of SD patients and the predictions of the semantic binding hypothesis. However, lexicality did have some effect, albeit small, on the occurrence of phoneme order errors in Gathercole et al.'s study.

Phonological errors in word recall, while rare in studies involving normal participants, are clearly of particular interest, as they make it possible to study the effect of semantic and lexical factors on the breakdown of phonological coherence in STM. Knott and Monsell (unpublished manuscript) found that healthy participants could be induced to make more frequent phoneme migration errors in their recall of words if they were presented with lists that contained an unpredictable mixture of words and nonwords. The nonword phonemes, which were presumably not tightly bound together as coherent items, appeared to damage the phonological integrity of the words. In these circumstances, the majority of recall errors were recombinations of phonemes from different list items, and clear effects of lexicality, frequency and imageability on the degree of phonological breakdown were observed. In summary, healthy participants' recall of mixed lists appeared to resemble the recall of pure word lists by SD patients.

The work presented in this chapter sought to replicate and extend these findings in order to address a number of issues. First, the mixed list methodology was used to explore the effects of lexicality, frequency and imageability on the degree of phonological breakdown in verbal STM. Participants were presented with lists in which phonemes were not repeated, allowing all phoneme migration errors to be traced. The effect of lexical and semantic variables on the occurrence of item and order errors was examined both at the level of whole items and individual phonemes. This allowed the predictions of the redintegration hypothesis (namely, that lexical/semantic factors primarily affect item and not order information) and the semantic binding hypothesis (lexical and semantic factors affect the occurrence of phoneme order as well as item errors) to be evaluated.

Secondly, the mixed list methodology made it possible to examine whether the lexical and semantic characteristics of the words had any impact on nonword recall. This issue is of particular theoretical importance because the semantic binding and redintegration accounts make rather different predictions. According to the semantic binding hypothesis, lexical and semantic constraints, which prevent the phonemes of more frequent and imageable words from breaking apart in STM, have a major impact on nonword recall in mixed lists, as the opportunity for nonword phonemes to migrate between list items is reduced. In contrast, the redintegration account apparently predicts that word frequency and imageability will affect the recall of words but not nonwords in mixed lists, as redintegration is thought to improve word recall at a late-stage, after the degree of phonological degradation has been determined for nonwords.

The work also compared the recall of words and nonwords in mixed lists (Experiment 1) with the same items in pure lists (Experiment 2), in order to explore the effect of mixing words with nonwords. Knott and Monsell's findings suggest that the presence of nonwords may have a detrimental impact on the recall of words in mixed lists. It is not yet clear, however, if the phonological stability of nonwords is improved by the presence of words. This issue is similar to the question of whether the lexical and semantic characteristics of the words in mixed lists affect nonword recall. Again, the item-based redintegration account apparently predicts no difference in nonword recall between mixed and pure lists, as the late-stage reconstruction of words should not improve the recall of nonwords. In contrast, the semantic binding account predicts that, in mixed lists, the phonological coherence of the words will improve the integrity of nonwords by reducing the extent to which nonword phonemes can migrate between list items.

Finally, five patients with SD were tested on these pure and mixed lists, allowing the similarities and differences in the recall errors of SD patients and healthy participants to be explored directly. If the errors of SD patients on pure word lists are qualitatively similar to the errors of healthy participants on mixed lists of words and nonwords, the suggestion that phoneme migration errors occur for the same reasons in these two populations/situations will be strengthened.

5.2 Experiment 1: Healthy participants tested on mixed lists of words and nonwords

In this experiment, normal participants were presented with five-item lists that contained an unpredictable mixture of words and nonwords. It was anticipated that the nonword phonemes would damage the phonological integrity of the words. In addition, it was predicted that lexicality, frequency and imageability would have a clear impact on the coherence of the phonological trace in these circumstances.

5.2.1 Method

5.2.1.1 Participants

The participants were 30 undergraduates, aged between 18 and 32, who spoke English as a first language and had normal hearing. They were tested individually and took part for course credit.

5.2.1.2 Design and materials

The experiment examined ISR for auditorily presented lists of five consonant-vowel-consonant (CVC) stimuli (see Appendix 10). Every list contained a mixture of words and nonwords. The proportion of words to nonwords was varied as a within-subjects factor. Lists could contain one word and four nonwords, two words and three nonwords or three words and two nonwords. Word frequency and imageability were also included as within-subjects factors. Words were assigned to four frequency by imageability groups on the basis of estimates of written word frequency and imageability taken from the Celex database (Baayen, Piepenbrock, & van Rijn, 1993) and the MRC psycholinguistic database (M. Coltheart, 1981). Mean frequency was 179 counts per million for the high frequency (HF) words (range = 51 – 656) and 6 counts per million for the low frequency (LF) words (range = 1 – 13). Mean imageability was 602 for the high imageability (HI) words (range = 573 – 659) and 442 for the low imageability (LI) words (range = 340 – 501). There were no significant frequency differences between groups that varied in imageability, and no significant imageability differences between groups that varied in frequency (all $t(58) < 1$). Words with homophones were excluded if the frequency or imageability of the homophone was higher than that of the target word. The words in each list were drawn from a single frequency and imageability group.

Each participant was tested on 60 lists. There were 15 lists in each of the frequency by imageability categories. For each word type, there were five lists for the three proportions of words to nonwords. The words/nonwords occurred in different serial positions across these five lists, in order to prevent the participants from anticipating which items would be words in advance. For lists containing one word (w) and four nonwords (n), the word occurred once in every serial position in the five lists (wnnnn, nwnnn, nnwnn, nnnwn, nnnnw). There were ten possible arrangements of words and nonwords for lists containing two or three words. Five were selected that minimised the degree to which the words and nonwords were clustered. In lists containing two words and three nonwords, the two words were never adjacent (wnwnn, wnnwn, nwnwn, nwnnw, nnwnw). In lists containing three words and two nonwords, the two nonwords were never adjacent (nwnww, nwwnw, wnwnw, wnwwn, wwnwn).

The nonwords were constructed from the words by recombining the initial consonants, vowels and final consonants to form new items. All the nonwords were legal and pronounceable. Lists were assembled so that phonemes were not repeated within a list. Items were not repeated in the course of the experiment.

5.2.1.3 Procedure

The items were recorded individually in a flat intonation by a female speaker and were digitised using a computer. Sound editing software (Cool Edit, Syntrillium) was used to position the items in the lists so that they occurred at a rate of one item per second. Presentation of the lists was controlled using SuperLab software (Cedrus). The order of the trials was re-randomised for each participant. A red exclamation mark appeared on the computer screen just prior to the start of each trial and remained until the list had finished playing. It was then replaced by a blue question mark, which prompted participants to try to recall the list aloud. Participants' responses were recorded on tape and were later transcribed. The participants were told in advance that the lists would contain a mixture of both words and nonwords and were given four practice trials. They were asked to recall the items in serial order and to make an attempt at each target, even if they were not completely sure they were correct.

5.2.2 Results

Omissions of items were positioned in the response transcripts in a way that minimised the numbers of phonemes occurring at the wrong serial positions (for example, if the items "ball, pid, wife..." were recalled as "bid, wife...", it was assumed that the first item was omitted and that 'bid' was recalled in the place of "pid"). In a small number of trials (0.6%), participants produced six rather than five items and the final response was discarded from the analysis.

The structure of the results section is as follows: first, overall recall accuracy is considered as a function of lexical and semantic factors and their interactions (i.e., word frequency, imageability, lexicality and the proportion of words to nonwords in the lists). Secondly, in an error analysis, the influence of lexical and semantic variables on memory for order and identity information is discussed at both the level of whole items and individual phonemes. Finally, order and identity errors are examined for each phoneme type separately (initial consonant – C1, vowel – V and final consonant – C2).

5.2.2.1 Recall accuracy

Table 5.1 shows the percentage of words and nonwords recalled in the correct serial position in each condition. These data were analysed using a within-subjects ANOVA. There were significant and substantial main effects of frequency ($F(1, 29) = 125.25$, $p < 0.0001$) and lexicality ($F(1, 29) = 244.12$, $p < 0.0001$), which occurred because recall was better for words than for nonwords and because items in lists containing high frequency words were recalled more accurately than items in lists containing low frequency words. The main effects of imageability ($F(1, 29) = 43.13$, $p < 0.0001$) and 'lexicality group', i.e., the proportion of words to nonwords in the list ($F(1, 29) = 10.55$, $p < 0.001$), also reached significance. Recall was better for lists containing high compared with low imageability words and recall improved when there were a larger number of words in the lists.

The imageability of the words in the lists had a bigger impact on the recall of words than nonwords ($F(1, 29) = 7.85$, $p < 0.01$). Bonferroni t tests indicated that the effect of imageability reached significance for both words ($t(29) = 6.12$, $p < 0.0001$) and nonwords ($t(29) = 4.97$, $p < 0.001$), but was larger for words. There was no interaction between word frequency and lexicality ($F(1, 29) < 1$), suggesting that word frequency not only enhanced the integrity of the words in STM but also had a substantial knock-on effect on nonword integrity. These results are consistent with the suggestion that the phonemes of highly frequent and imageable words are bound together more strongly, reducing the opportunity for nonword phonemes to migrate.

Table 5.1: *Percentage of words and nonwords recalled in mixed lists (Experiment 1) as a function of frequency, imageability and the proportion of words to nonwords*

	Words: nonwords in list	Words M	Words SD	Nonwords M	Nonwords SD
High frequency, high imageability	3:2	77.8	12.7	48.0	19.9
	2:3	66.7	18.1	51.1	22.0
	1:4	66.7	23.7	45.5	20.9
	Mean	70.4	14.5	48.2	17.9
High frequency, low imageability	3:2	75.1	14.2	48.7	24.5
	2:3	57.3	19.5	36.7	17.4
	1:4	58.0	20.6	40.7	16.6
	Mean	63.5	14.3	42.0	16.4
Low frequency, high imageability	3:2	55.3	17.7	31.7	19.1
	2:3	58.3	16.4	38.2	18.5
	1:4	66.7	25.4	39.7	19.0
	Mean	60.1	15.0	36.5	14.8
Low frequency, low imageability	3:2	58.2	19.1	31.3	18.7
	2:3	41.3	19.3	31.3	16.0
	1:4	42.0	20.6	31.2	15.6
	Mean	47.2	15.3	31.3	14.2
Mean	3:2	66.6	11.9	39.9	15.9
	2:3	55.9	13.4	39.3	15.4
	1:4	58.3	16.0	39.3	15.4
	Grand Mean	60.3	12.7	39.5	14.7

There was also a significant interaction between the lexical status of items as words or nonwords and the proportion of words to nonwords in the lists ($F(2, 58) = 11.53$, $p < 0.0001$). Bonferroni t tests indicated that the word items were recalled more accurately when there were fewer nonwords in the lists (2 words vs. 3 words: $t(29) = 7.15$, $p < 0.0001$, 1 word vs. 2 words: $t(29) = 1.24$, n.s.). In contrast, the proportion of words to nonwords had no effect on recall accuracy for nonwords (1 word vs. 2 words: $t(29) < 1$ and 2 words vs. 3 words: $t(29) < 1$). This result is rather surprising, given that word frequency and imageability had an impact on nonword recall. However, the proportion of words to nonwords was found to influence the phonological coherence of nonwords in the error analysis below.

There was a significant frequency by lexicality group interaction ($F(2, 58) = 6.56$, $p < 0.01$), as frequency had a bigger impact on recall when there were more words in the lists. Bonferroni t tests indicated that the effect of frequency was significant for lists containing one word ($t(29) = 3.82$, p < 0.01), two words ($t(29) = 5.85$, $p < 0.0001$) and three words ($t(29) = 8.23$, $p < 0.0001$), although the effect was largest for lists containing three words.

In contrast, imageability had a bigger impact on recall when there were fewer words in the lists ($F(2, 58) = 11.70$, $p < 0.0001$). Bonferroni t tests indicated that there was a significant imageability effect for lists containing one or two words ($t(29) = 5.18$, $p < 0.001$ and $t(29) = 6.02$, $p < 0.0001$) but not for lists containing three words ($t(29) = 1.04$, n.s.). Imageability may have played a greater role in ISR when the coherence of the phonological trace was jeopardised by the presence of a large number of nonwords in the lists.

Finally, the four-way interaction between frequency, imageability, lexicality group and lexicality reached significance ($F(2, 58) = 3.28$, $p < 0.05$). This was accounted for by the fact that there was a three-way interaction between frequency, imageability and lexicality group for words ($F(2, 58) = 3.45$, $p < 0.05$) but not for nonwords ($F(2, 58) = 1.70$, n.s.). Imageability, therefore, made a greater contribution to word recall when there were more nonwords in the list, particularly when the words were also low in frequency. The imageability by lexicality group interaction in word recall reached significance when frequency was low ($F(2, 58) = 11.04$, $p < 0.0001$) but not when frequency was higher ($F(2, 58) < 1$), consistent with the suggestion that imageability effects were larger when the coherence of items in STM was jeopardised by other factors (i.e., a higher proportion of nonwords in the list or low word frequency).

The influence of these lexical and semantic factors on recall accuracy was also examined in a by-items analysis, in which lexicality, frequency, imageability and the proportion of words to nonwords were entered as between-cases factors. The main effects of lexicality ($F(1, 276) = 64.59$, $p < 0.0001$), frequency ($F(1, 276) = 22.40$, $p < 0.0001$) and imageability ($F(1, 276) = 9.13$, $p < 0.01$) were significant as before, but the proportion of words to nonwords did not have a significant impact on overall recall accuracy ($F(2, 276) = 2.01$, n.s.). The interaction between imageability and lexicality group approached significance ($F(2, 276) = 2.85$, $p = 0.08$), consistent with the previous finding that imageability had a larger influence when there were more nonwords in the list. No other interactions reached significance. As in the by-subjects analysis, word frequency and imageability influenced the recall of both words and nonwords. Independent-samples t tests indicated that frequency affected both word recall ($t(118) = 4.14$, $p < 0.01$) and nonword recall ($t(178) = 3.29$, $p < 0.01$). Similarly, the effect of imageability approached significance for words ($t(118) = 1.83$, $p = 0.07$) and reached significance for nonwords ($t(178) = 2.05$, $p < 0.05$).

5.2.2.2 Error analysis

The influence of lexical and semantic factors (lexicality, frequency, imageability and the proportion of words to nonwords) on the number of order and identity errors was examined at both the level of whole items and individual phonemes. Order errors occurred when an item or a phoneme was recalled correctly but in the wrong serial position. Identity errors included both omissions and commission errors (i.e., intrusions of incorrect phonemes from outside the list) and occurred when an item or phoneme was not recalled correctly at any serial position.

Lexicality had opposite effects on identity and order errors at the level of whole items. Figures 5.1a and 5.1b show the mean number of whole-item order and identity errors for words and nonwords as a percentage of the number of items presented. The data from mixed lists of words and nonwords (Experiment 1) are shown in conjunction with the data examining pure lists (Experiment 2), although only the Experiment 1 data are discussed here. There was a highly significant effect of lexicality on the number of item identity errors ($t(29) = 17.34, p < 0.0001$), as these errors occurred much more frequently for nonwords than for words. In contrast, whole item transpositions occurred more commonly for words than for nonwords ($t(29) = 5.70, p < 0.0001$).

Figures 5.2a and 5.2b show the mean number of order and identity errors at the level of individual phonemes for words and nonwords. Identity errors at the phoneme level, like those at the whole-item level, occurred less frequently for words than for nonwords ($t(29) = 9.99$, $p < 0.0001$). Phoneme order errors also occurred less frequently for words than for nonwords ($t(29) = 3.65$, $p < 0.01$), in contrast with order errors at the whole-item level. This result suggests that nonwords were more likely to fragment in verbal STM. Whole item transpositions were more common for words than for nonwords, but individual phonemes were more likely to migrate for nonwords.

Frequency affected the occurrence of order and identity errors in a similar way to lexicality. Identity errors at the whole item level were more common for low compared with high frequency words (high frequency mean = 27.1% of items presented, SD = 10.7; low frequency mean = 42.9%, SD = 12.9; $t(29) = 10.11, p < 0.0001$). In contrast, there was no significant difference in the rate of item order errors between high and low frequency words (high frequency mean = 4.4% of items presented, SD = 3.7; low frequency mean = 3.3%, SD = 2.9; $t(29) = 1.84$, n.s.).

Figures 5.3a and 5.3b show order and identity errors at the level of individual phonemes for high and low frequency words, and the nonwords they were mixed with, as a percentage of the number of phonemes presented. Identity errors at the phoneme level, like those at the whole-item level, occurred less frequently for high than low frequency words ($t(29) = 7.63$, $p < 0.0001$). Phoneme order errors also occurred less frequently for high than low frequency words ($t(29) = 2.11$, $p < 0.05$), in contrast with order errors at the whole-item level. As the larger number of phoneme order errors for low frequency words could not be accounted for by whole-item migrations, this finding suggests that frequency affected the extent to which the phonemes of words were recalled together as a coherent item.

Interestingly, word frequency affected the occurrence of phoneme order and identity errors for nonwords as well as for words in mixed lists (see Figures 5.3a and 5.3b). Phoneme order errors were less common for nonwords that had been presented with high compared with low frequency words ($t(29) = 4.01$, $p < 0.001$), as were phoneme identity errors ($t(29) = 6.10$, $p < 0.0001$). These results are consistent with the suggestion that the phonemes of high frequency words were more likely to remain together in verbal STM, reducing the opportunity for nonword phonemes to migrate in mixed lists. Word frequency appeared to impinge on the stability of the whole phonological trace, and not just on the phonological representation of the word items.

Imageability had a less sizeable impact than frequency on phoneme identity and order errors, presumably because this factor had a rather smaller effect on overall recall accuracy. Nevertheless, several of the word frequency findings were replicated for imageability. Figures 5.4a and 5.4b show identity and order errors at the level of individual phonemes for high and low imageability words and the nonwords they were mixed with. Phoneme identity errors were less common for high than low imageability words ($t(29) = 5.52$, $p < 0.0001$), although imageability did not affect the number of phoneme migration errors ($t(29) = 1.18$, n.s.). Again, the imageability of the words impinged on the stability of the phonological trace for nonwords. Phoneme order errors were less common for nonwords presented with high compared with low imageability words ($t(29) = 2.14$, $p < 0.05$), as were phoneme identity errors ($t(29) = 5.18$, $p < 0.0001$).

The effect of lexicality group (i.e., the proportion of words to nonwords in the lists) on the number of order and identity errors was also examined at the level of individual phonemes (see Figures 5.5a and 5.5b). There were fewer phoneme migration errors when the proportion of words to nonwords was higher, and this effect occurred for both words (1 vs. 2 words in list; $t(29) < 1$; 2 vs. 3 words in list; $t(29) = 3.54, p < 0.01$) and nonwords (1 vs. 2 words in list; $t(29) < 1$; 2 vs. 3 words in list; $t(29) = 3.68, p < 0.001$). In contrast, the proportion of words to nonwords in the lists only influenced the number of phoneme identity errors for words (1 vs. 2 words in list; $t(29) < 1$; 2 vs. 3 words; $t(29) = 3.49, p < 0.01$) and had no effect on nonwords (1 vs. 2 words in list; $t(29) < 1$; 2 vs. 3 words; $t(29) = 1.81$, n.s.). These results are again consistent with the suggestion that word phonemes are bound together more than nonword phonemes in verbal STM, reducing the opportunity for nonword phonemes to migrate in lists containing a higher proportion of words.

Analyses examining the number of order and identity errors for C1, V and C2 phonemes as a function of lexicality were also conducted (see Figures 5.6a and 5.6b). There was a main effect of phoneme type for both order errors ($F(2, 58) = 118.79, p < 0.0001$) and identity errors ($F(2, 58) = 45.41, p < 0.0001$), suggesting that vowels were recalled more accurately than consonants. For phoneme migration errors, there was no interaction between phoneme type and lexicality ($F(2, 58) < 1$), indicating that the advantage for vowels was of the same magnitude for words and nonwords (collapsing across lexicality, V vs. C1: Bonferroni $t(29) = 12.59, p < 0.0001$; V vs. C2: Bonferroni $t(29) = 15.29, p < 0.0001$). This finding suggests that the greater phonological stability of the words helped to prevent migrations of both vowels and consonants.

In contrast, the phoneme identity error difference between vowels and consonants was smaller for words than for nonwords ($F(2, 58) = 31.08, p < 0.0001$). Bonferroni t tests revealed that for nonwords, fewer identity errors involved V than C1 phonemes ($t(29) = 3.06, p < 0.01$) and V than C2 phonemes ($t(29) = 8.89, p < 0.0001$), whereas for words, there was no difference between V and C1 phonemes ($t(29) = 1.09$, n.s.) and a less marked difference between V and C2 phonemes ($t(29) = 3.86, p < 0.01$). The acoustic energy of the vowels may have assisted their identification and maintenance, minimising the number of vowel identity errors. This effect would have been less critical for words than for nonwords, however, since identification and memory of each word phoneme would have been supported by knowledge about the whole word.

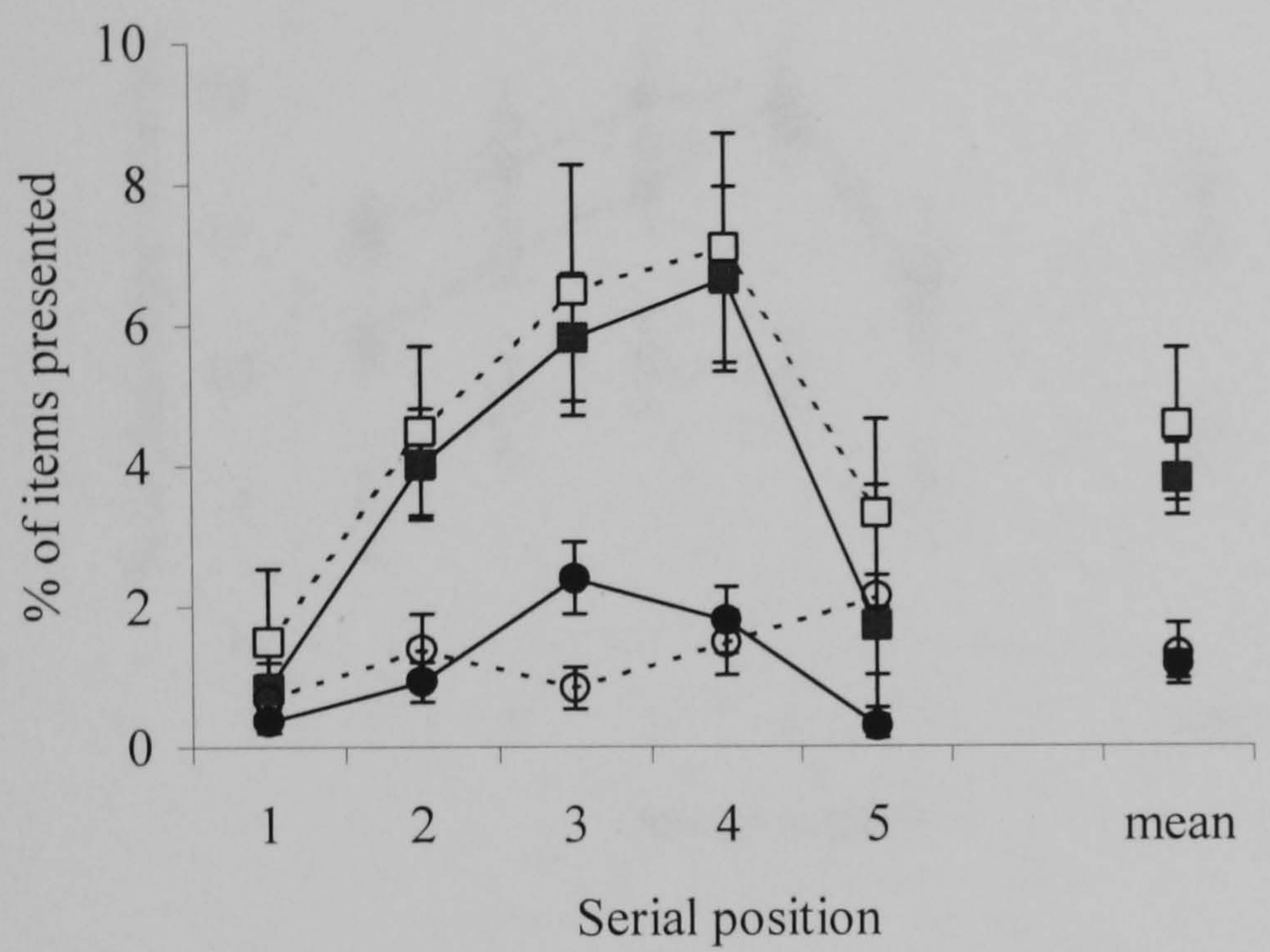
5.2.3 Discussion

In this experiment in which participants attempted to recall lists composed of a mixture of words and nonwords, considerable numbers of phoneme order and identity errors occurred for words as well as nonwords. In contrast, phonological errors are relatively uncommon in studies involving the recall of pure word lists (e.g., Henson et al., 1996; Pickering et al., 1998). It appears, therefore, that mixing words with nonwords impaired the phonological coherence of the words in this study, possibly because the nonword phonemes, which were not tightly bound together as coherent items, were able to recombine with the word phonemes. Some specific support for this suggestion was obtained in analyses that examined the impact of manipulating the proportion of words to nonwords in the lists. Word recall declined as the number of nonwords was increased because the phonemes of words were more likely to migrate between list items or be recalled incorrectly when the proportion of nonwords was larger.

This mixed-list methodology made it possible to examine the impact of a number of lexical and semantic variables on the stability of the phonological trace. There were clear effects of lexicality, word frequency and imageability, as the words were recalled more accurately than the nonwords and recall was better for high frequency/imageability words compared with low frequency/imageability words. There were higher order interactions between these variables that suggested that semantic constraints played a larger role in verbal STM when the phonological coherence of items was jeopardized for other reasons, for example, when the proportion of nonwords in the lists was large.

Figure 5.1: *Order and identity errors at the level of whole items, as a function of lexicality (Experiments 1 and 2)*

Fig. 5.1a: Order errors



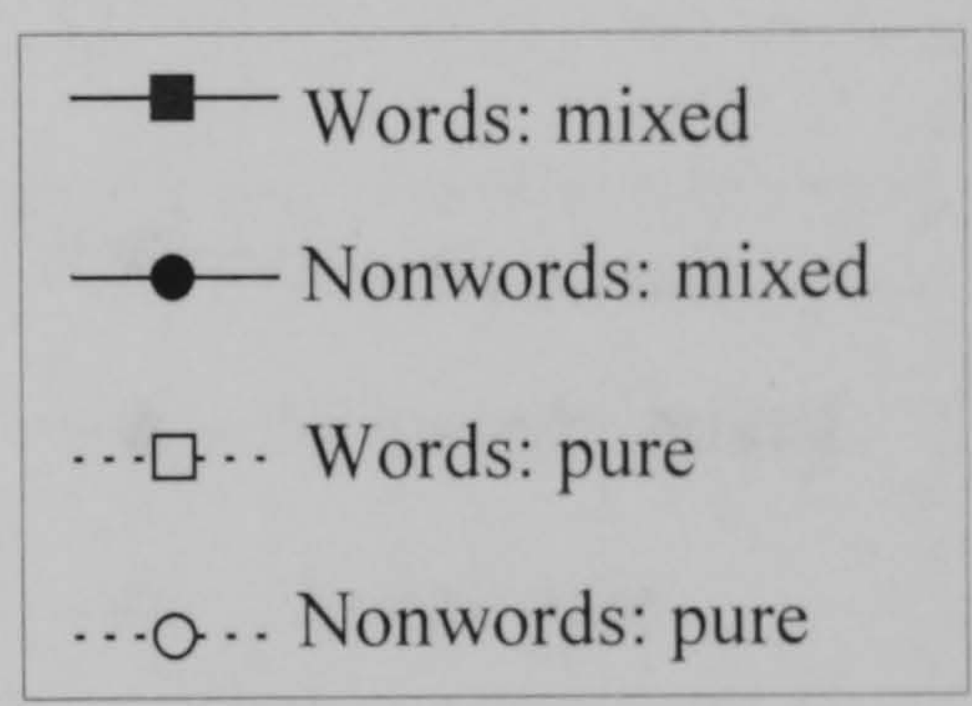
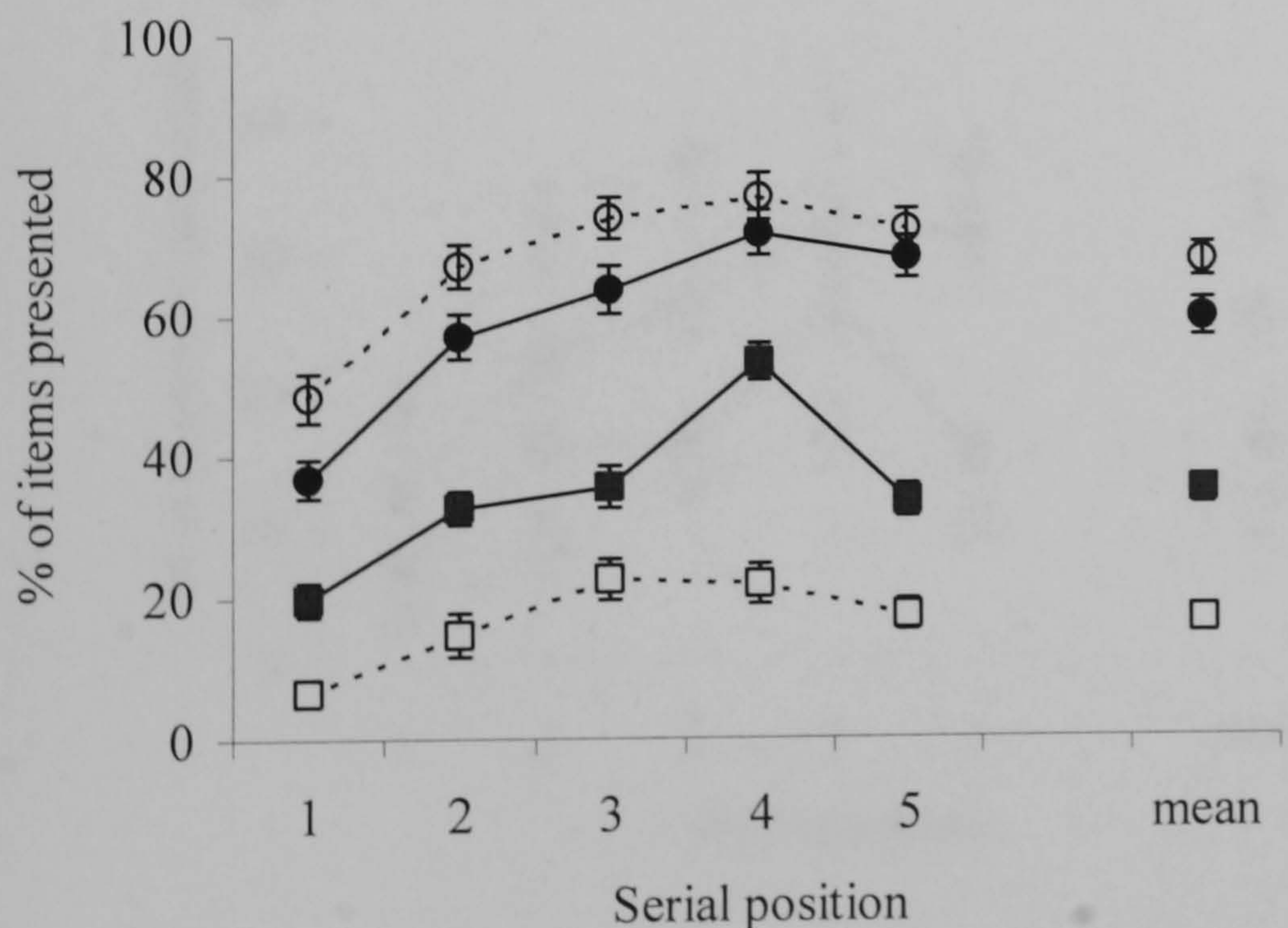


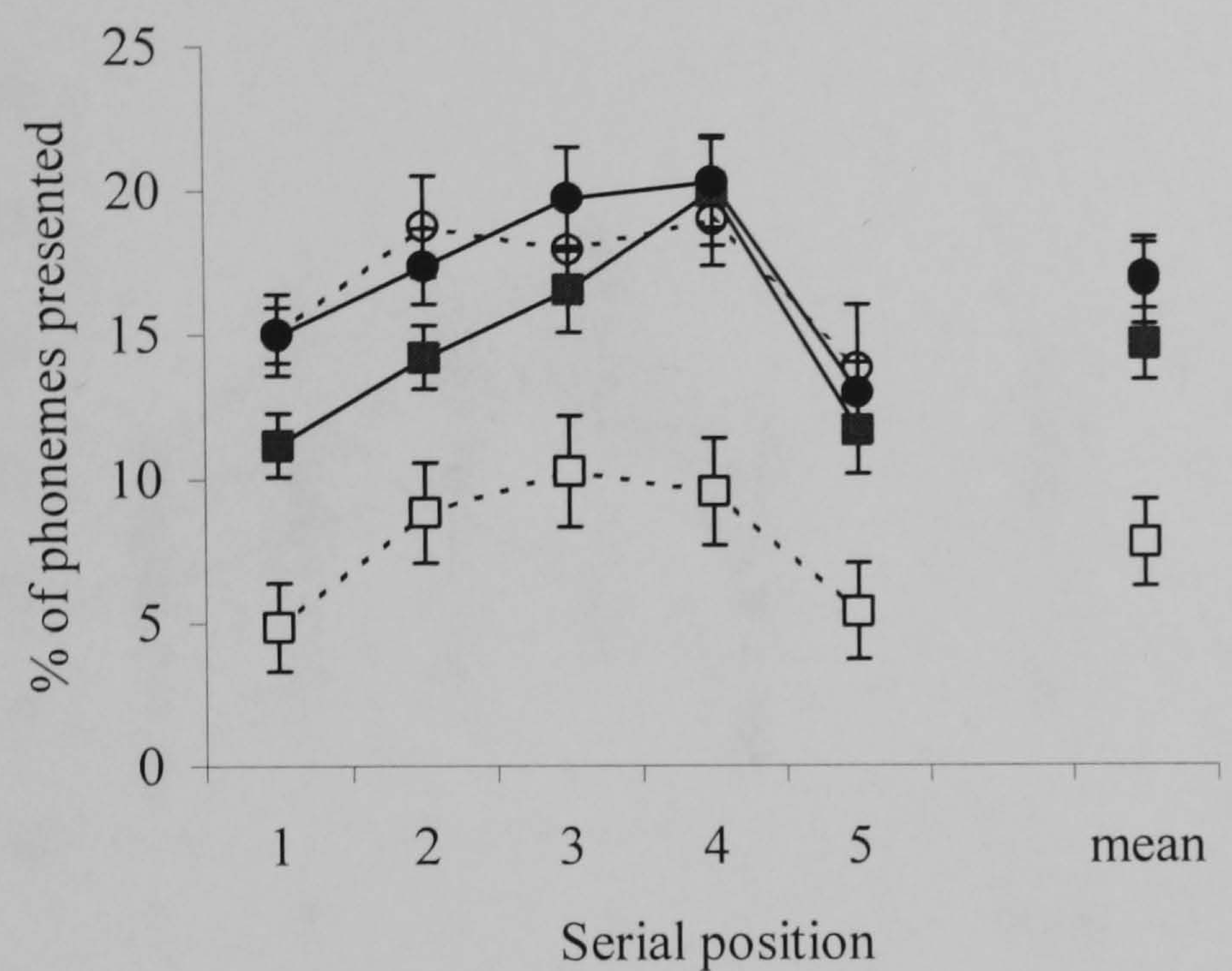
Fig. 5.1b: Identity errors



Error bars show SE

Figure 5.2: *Order and identity errors at the level of individual phonemes, as a function of lexicality (Experiments 1 and 2)*

Fig. 5.2a: Order errors



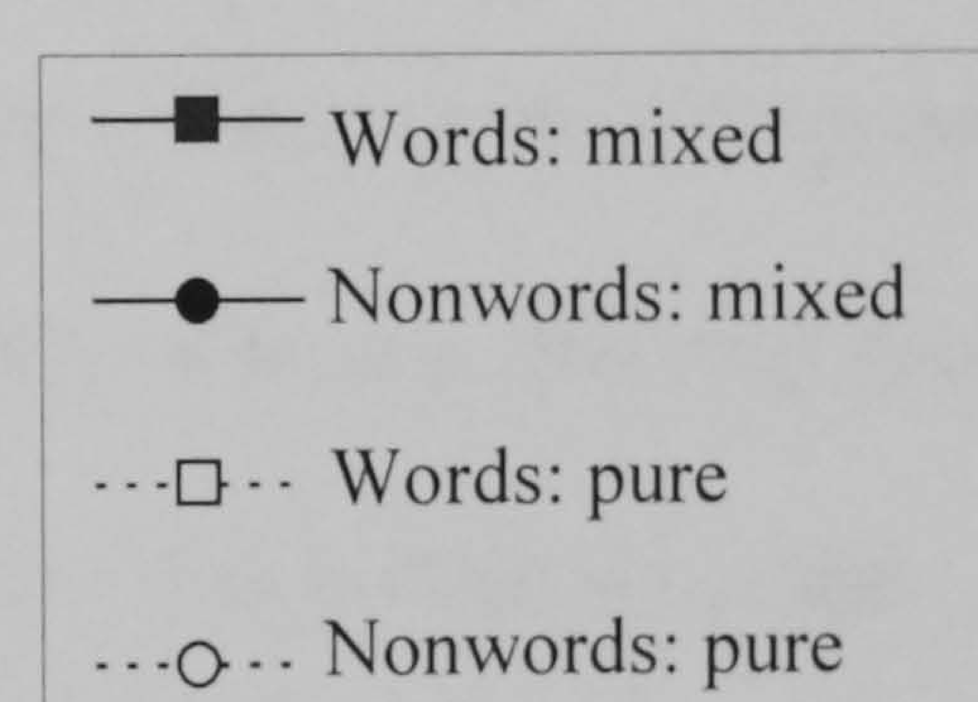
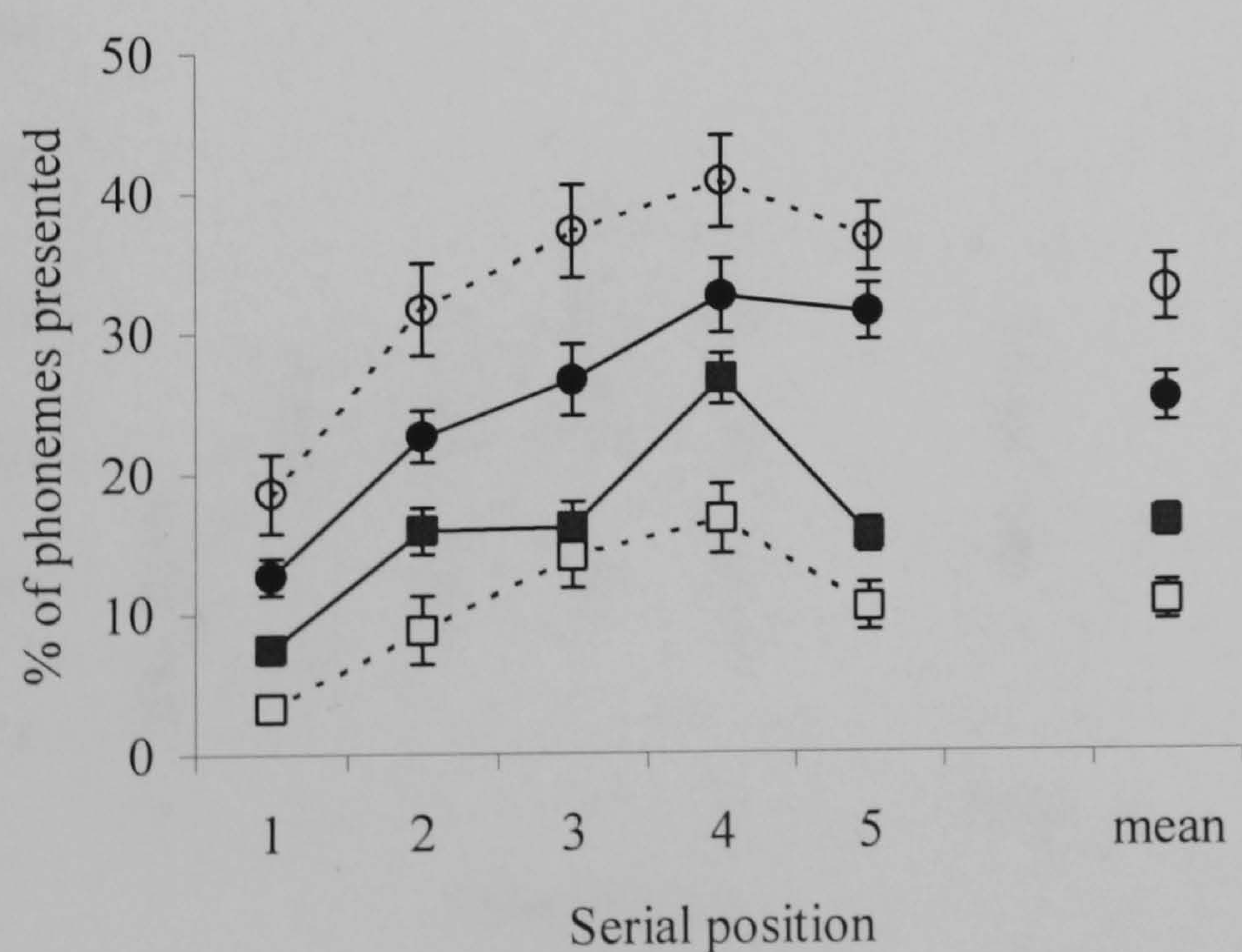


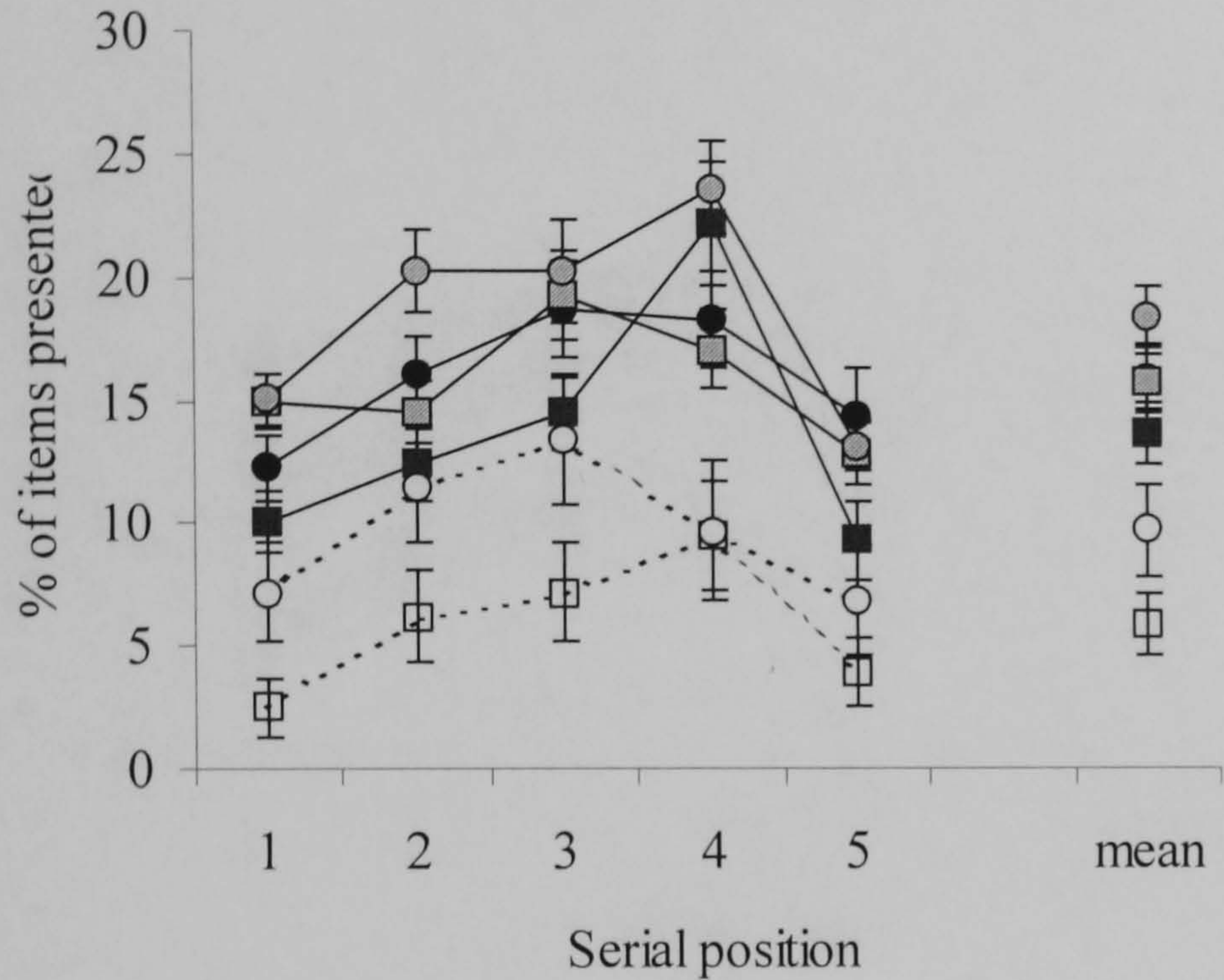
Fig. 5.2b: Identity errors



Error bars show SE

Figure 5.3: *Order and identity errors at the level of individual phonemes, as a function of frequency (Experiments 1 and 2)*

Fig. 5.3a: Order errors



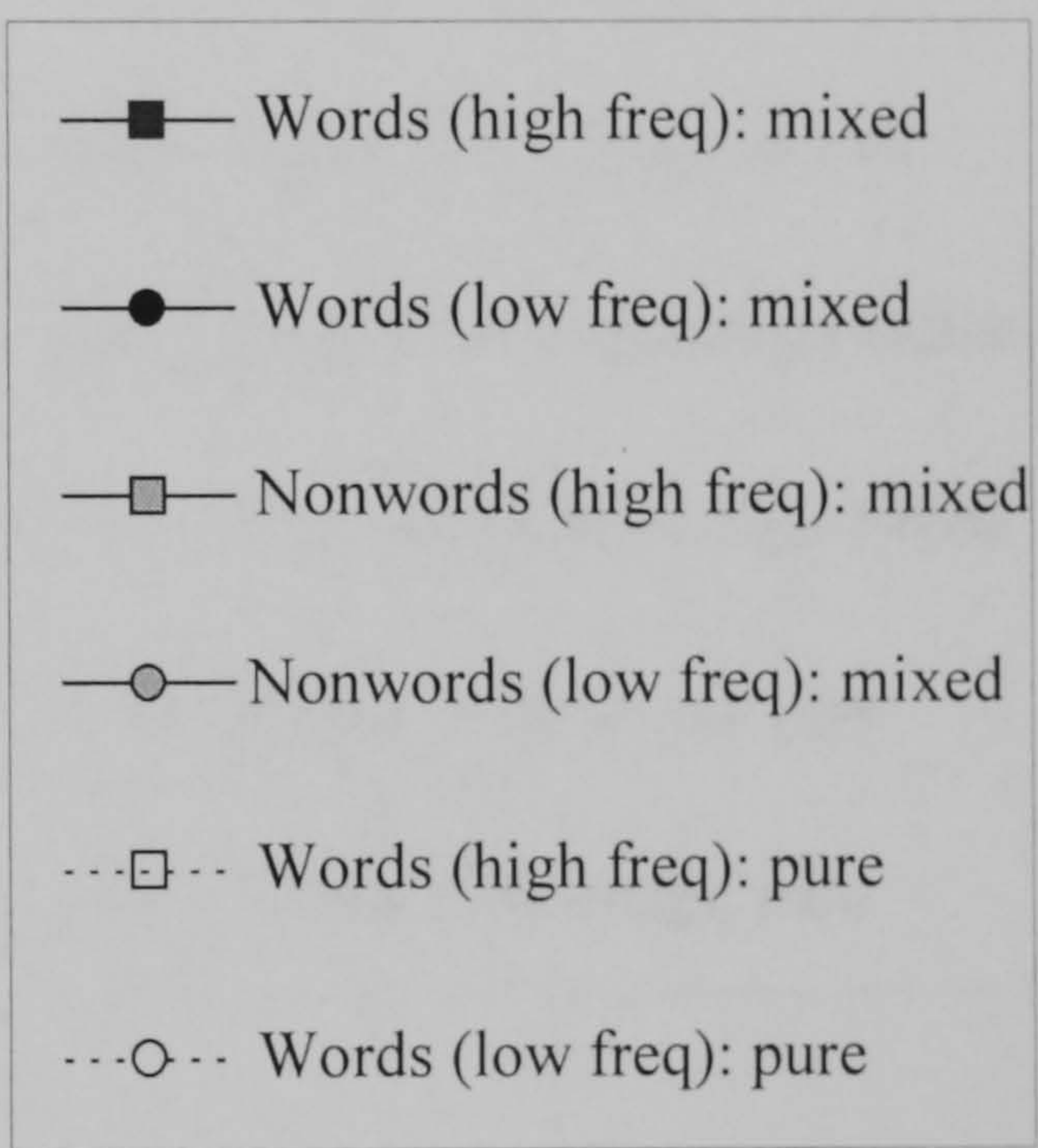
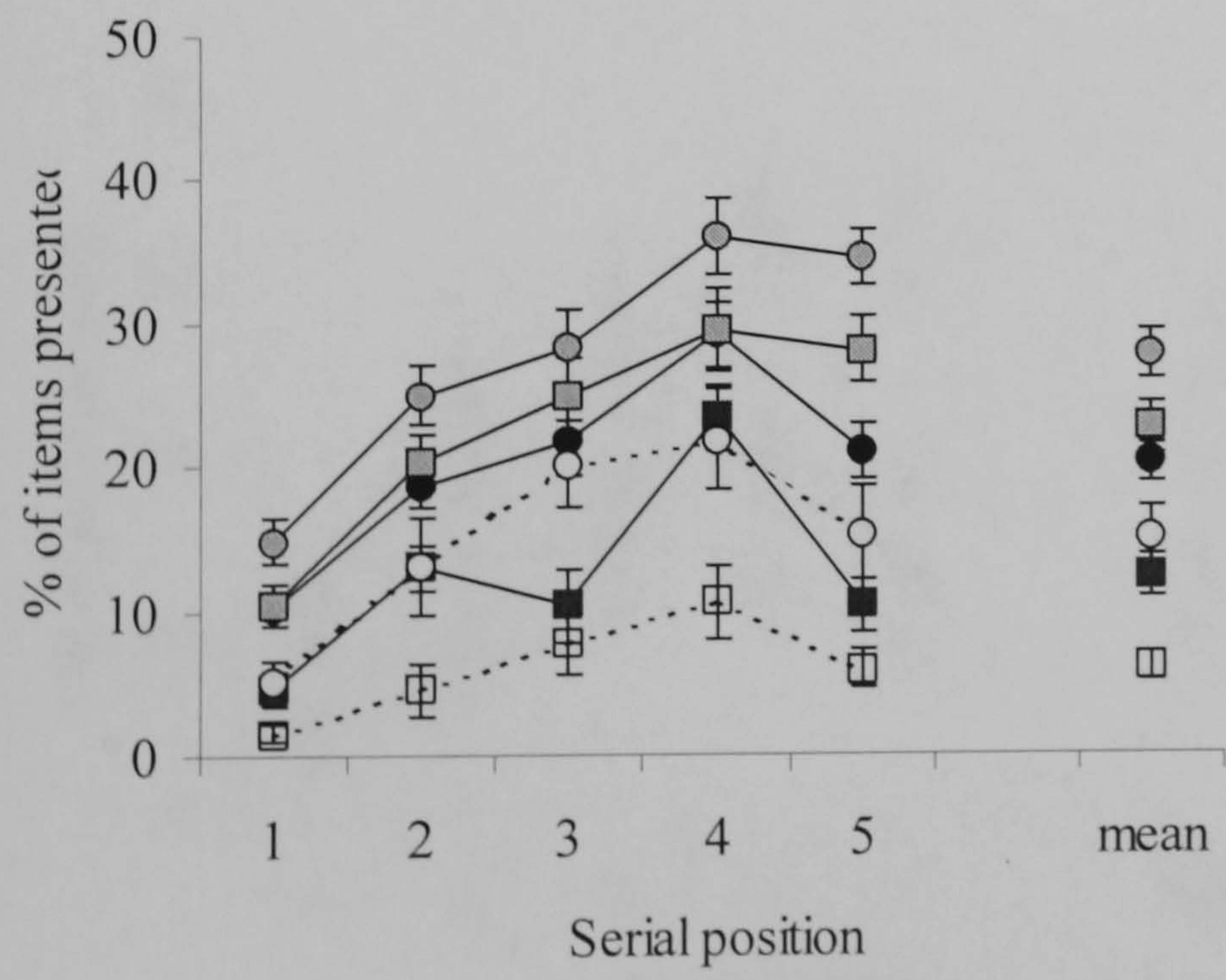


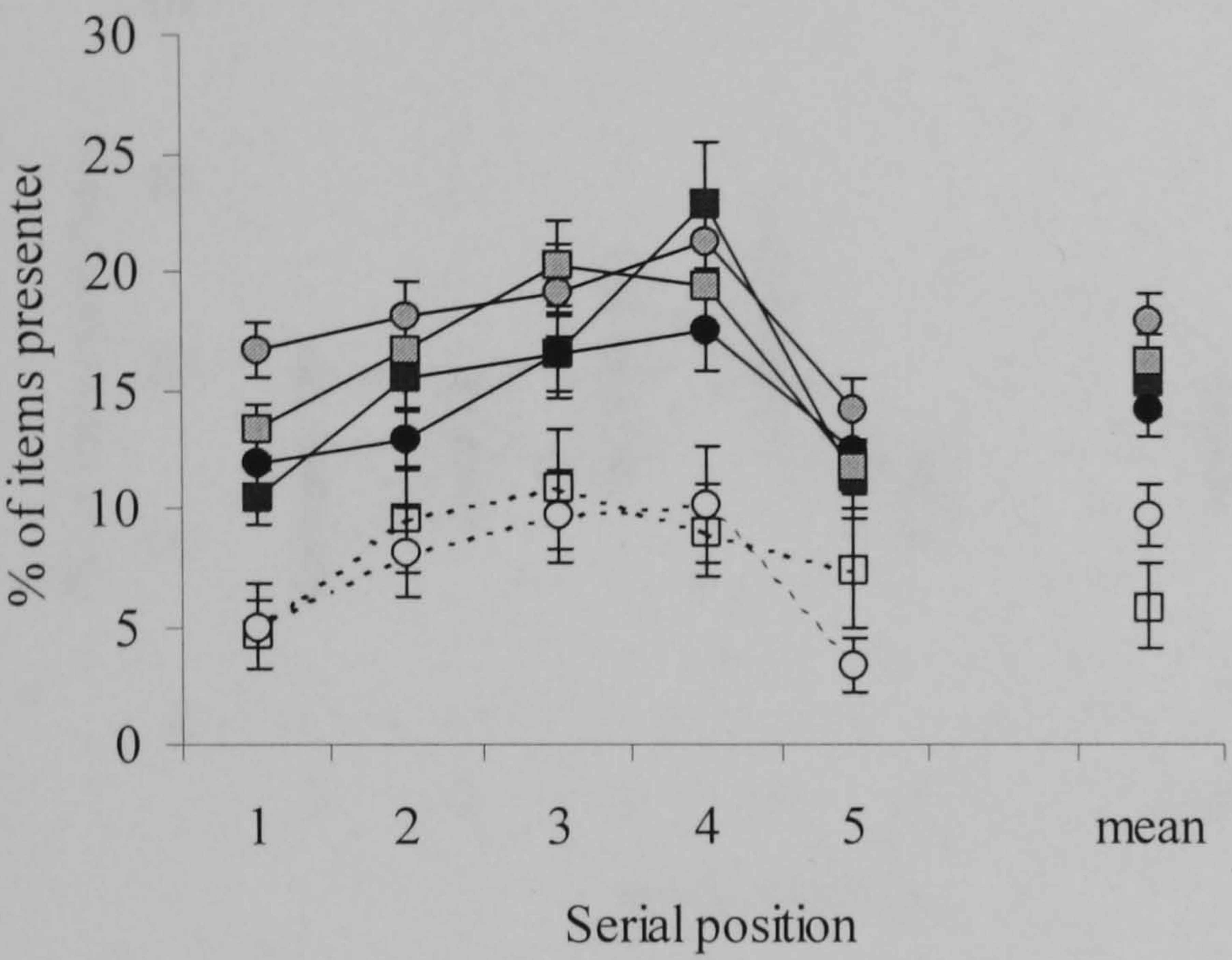
Fig. 5.3b: Identity errors



Error bars show SE

Figure 5.4: *Order and identity errors at the level of individual phonemes, as a function of imageability (Experiments 1 and 2)*

Fig. 5.4a: Order errors



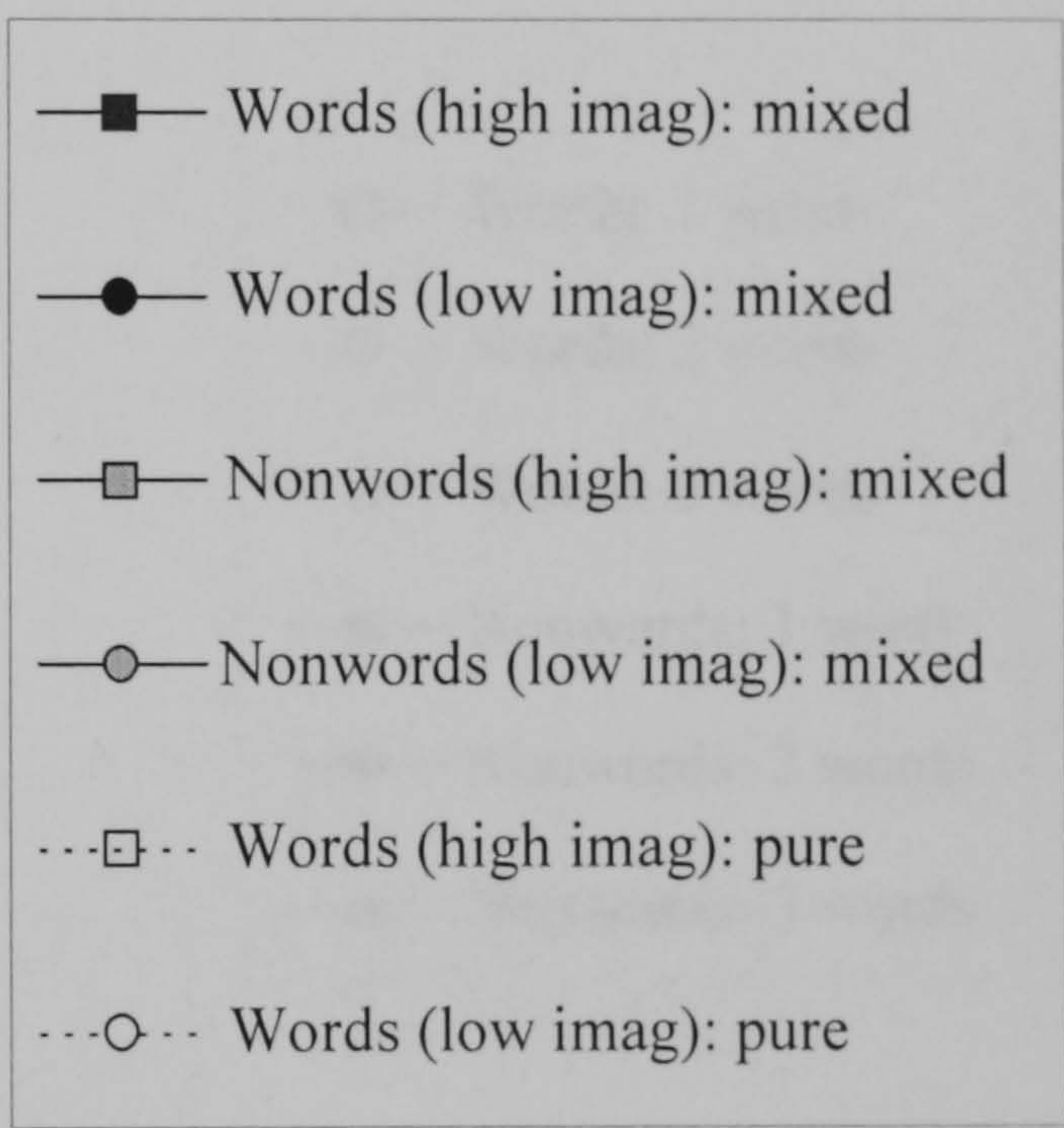
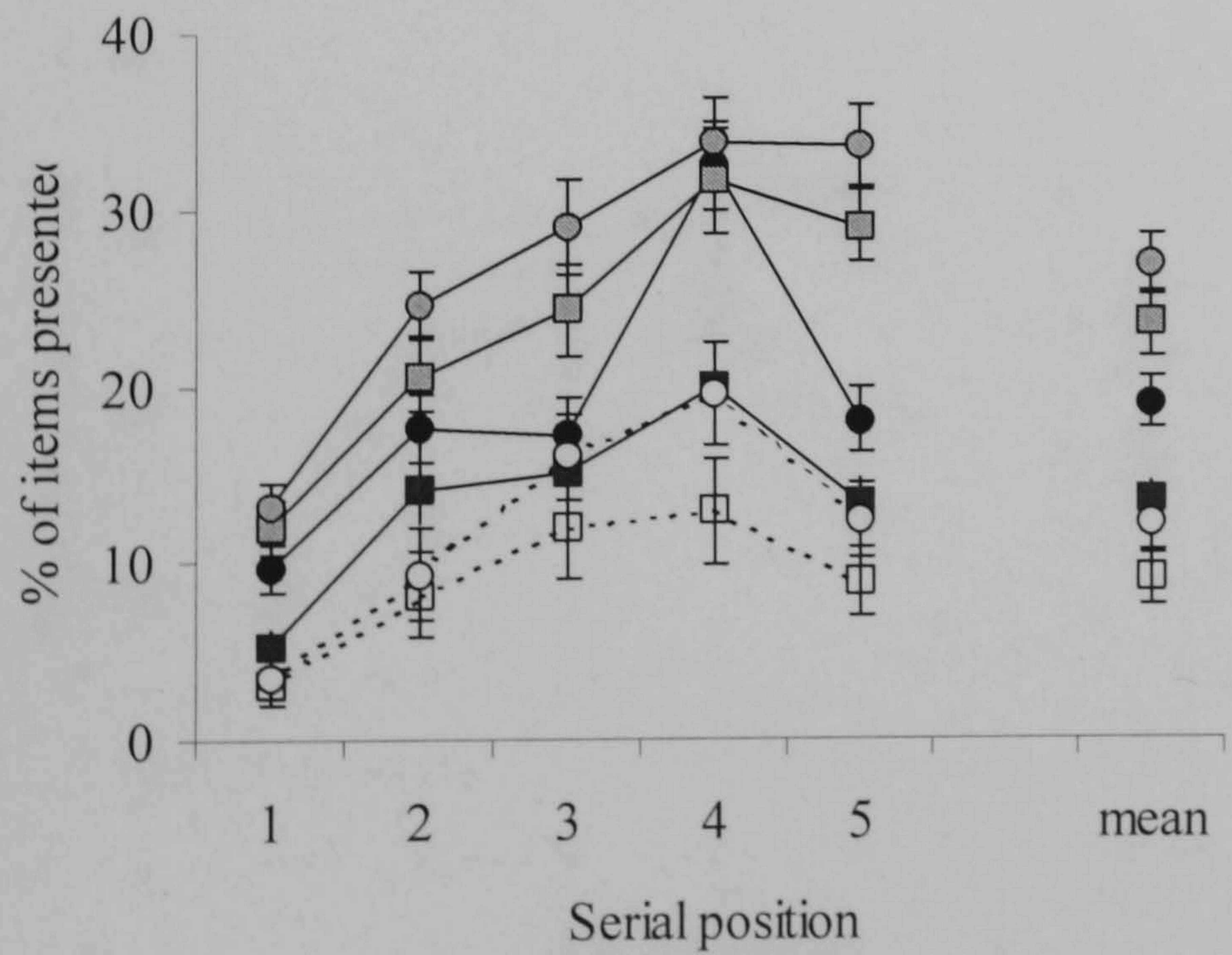


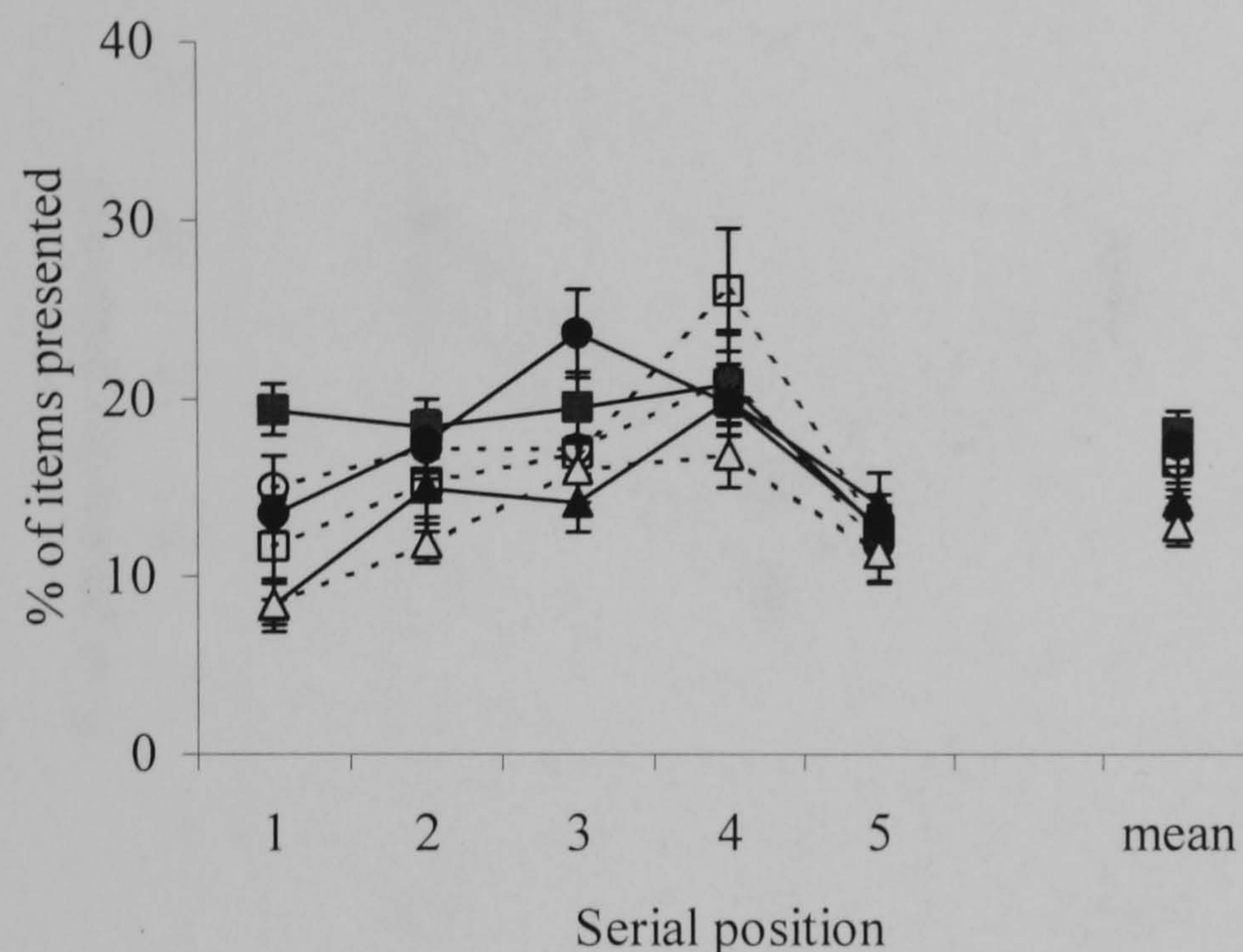
Fig. 5.4b: Identity errors



Error bars show SE

Figure 5.5: *Order and identity errors at the level of individual phonemes, as a function of the number of words in mixed lists (Experiment 1)*

Fig. 5.5a: Order errors



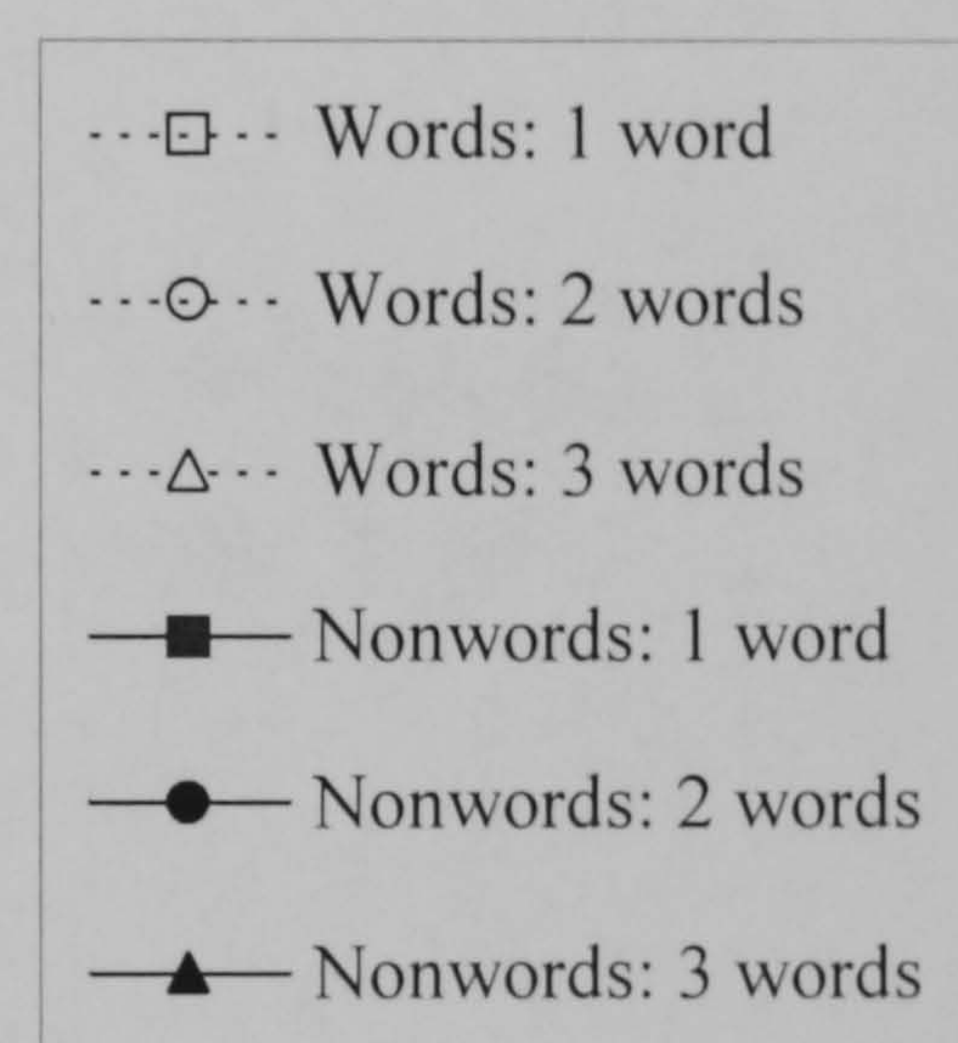
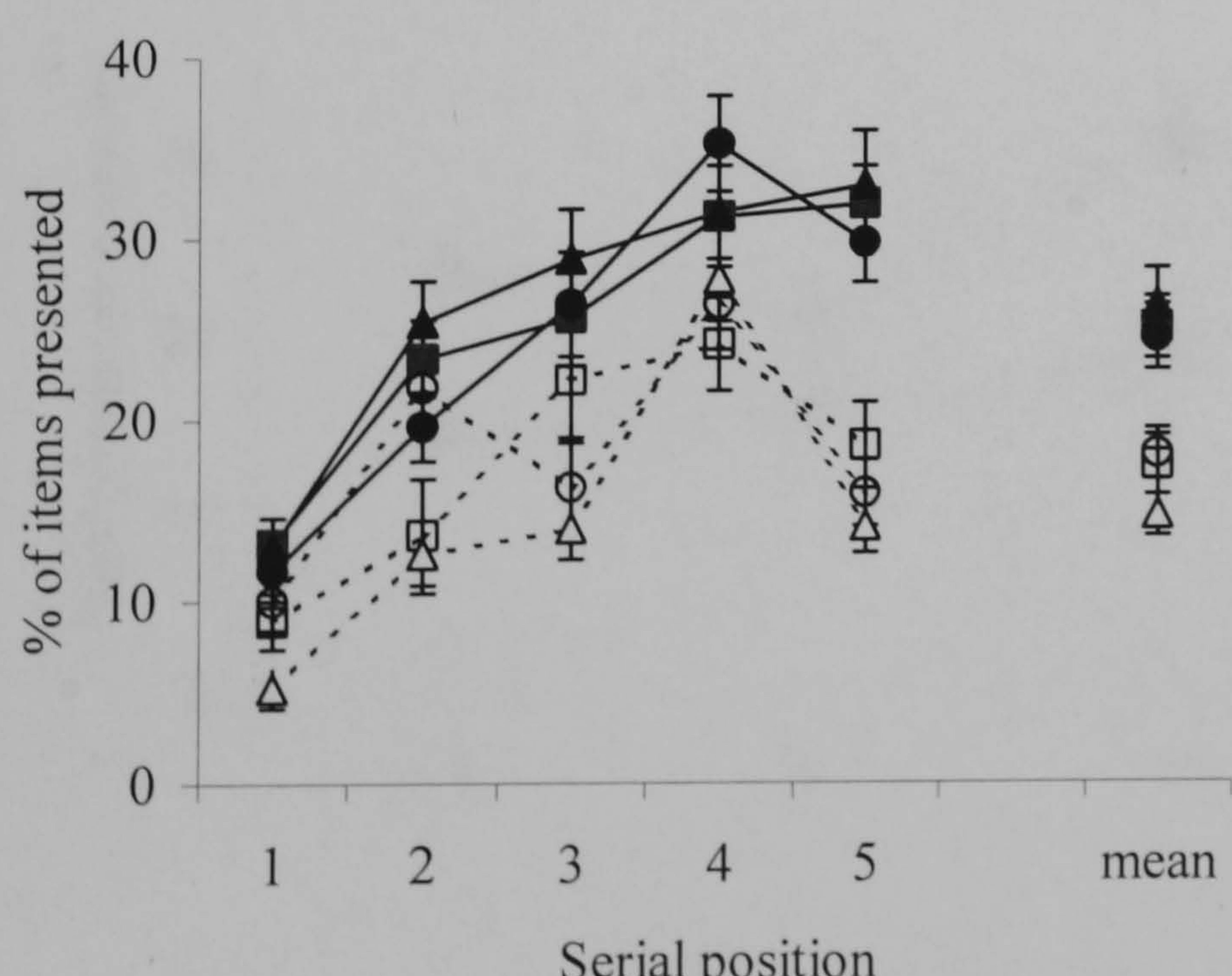


Fig. 5.5b: Identity errors



Error bars show SE

Figure 5.6: *Order and identity errors at the level of individual phonemes, as a function of phoneme type and lexicality in mixed lists (Experiments 1)*

Fig. 5.6a: Order errors

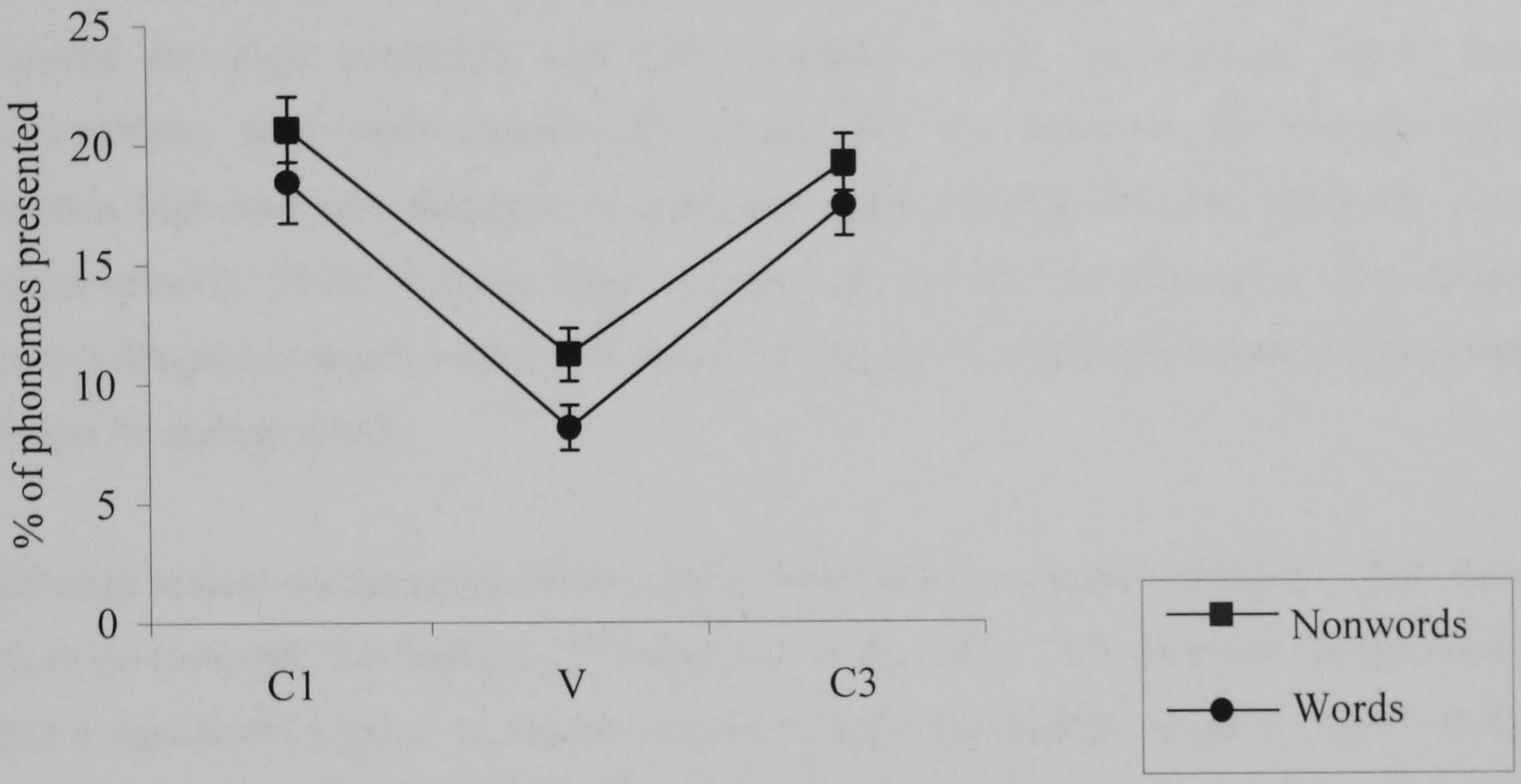
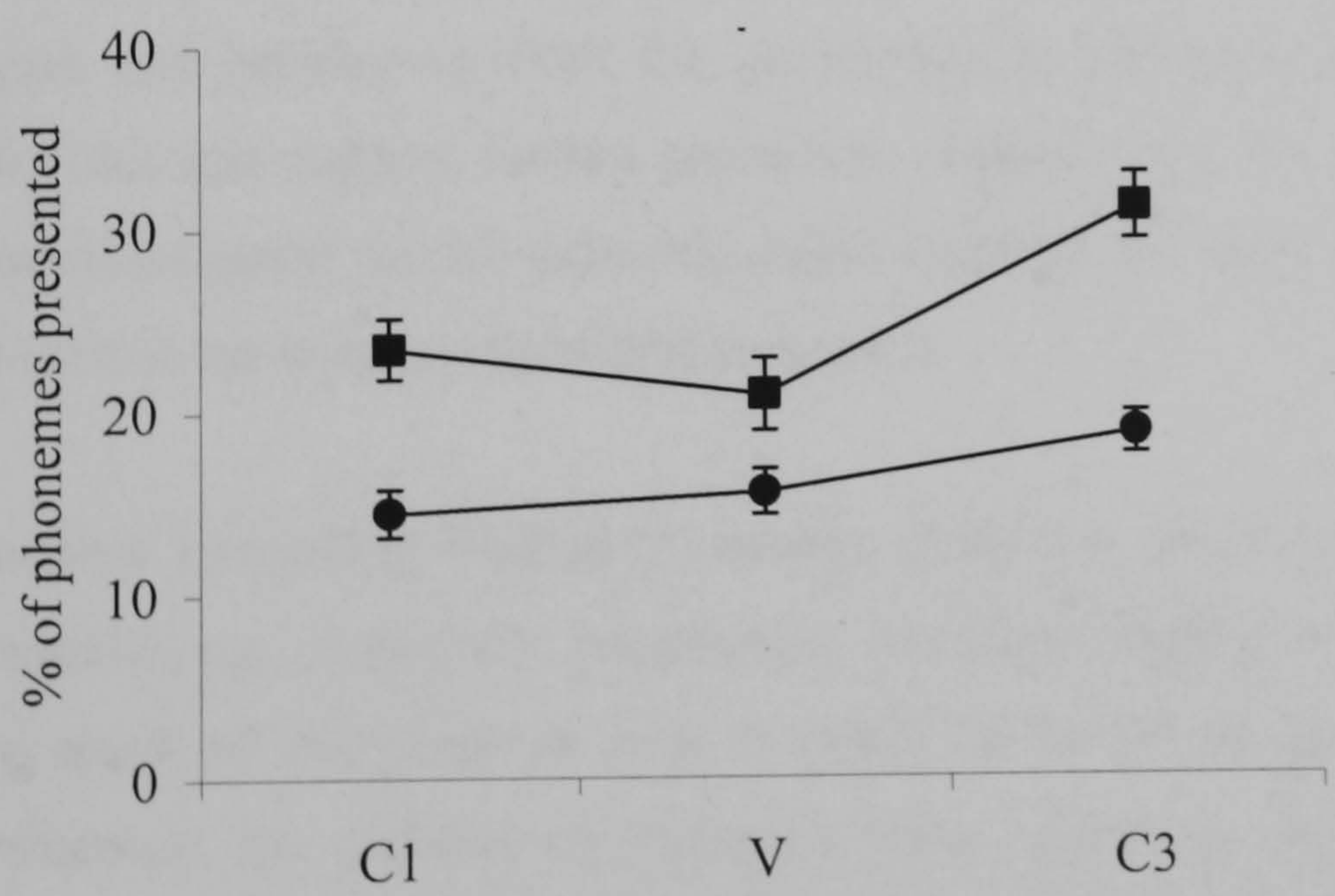


Fig. 5.6b: Identity errors



Error bars show SE

Lexical and semantic factors affected both identity and order errors at the level of individual phonemes. Phoneme identity errors were less common for words compared with nonwords, and were more infrequent when word frequency and imageability were high, suggesting that lexical and semantic knowledge constrained memory for phoneme identity. In addition, the phonemes of words were less likely to migrate to new list positions than the phonemes of nonwords and similarly, fewer phoneme migration errors occurred for high compared with low frequency words. In contrast, whole item transpositions were more common for words than for nonwords and did not differ between high and low frequency words (see Saint-Aubin & Poirier, 2000, for some similar results). These findings taken together suggest that the phonemes of nonwords and low frequency words were more likely to fragment in verbal STM than the phonemes of high frequency words.

Although lexical and semantic factors had a larger effect on phoneme identity than order errors (in line with the findings of Gathercole et al., 2001), lexicality and frequency did have a significant impact on the occurrence of phoneme migration errors in this study, supporting the predictions of the semantic binding hypothesis (outlined in Section 5.1). These results are consistent with the claim that the phonemes of familiar words are more strongly associated in the phonological system, facilitating their binding into coherent items and helping to limit the occurrence of phoneme migration errors. This study provides less support for the prediction, derived from the redintegration approach, that lexical/semantic factors primarily affect memory for item rather than order information, at least at the level of individual phonemes.

Another interesting finding to emerge from this study was that lexical and semantic variables, e.g., frequency, imageability and the proportion of words to nonwords, affected the recall of nonwords as well as words in the mixed lists. All three of these factors influenced the number of phoneme order errors for nonwords, consistent with the suggestion that the phonemes of frequent and imageable words were more bound together, reducing the opportunity for nonword phonemes to migrate between list items. Frequency and imageability also reduced the number of phoneme identity errors for nonwords, suggesting that the coherence of the word items had a knock-on effect on the integrity of the whole phonological trace. These findings are consistent with accounts that view phonological STM as being underpinned by language processing mechanisms that incorporate lexical and semantic constraints (e.g., the semantic binding hypothesis), as the ongoing effect of these constraints will impact on the stability of the phonological representation for nonwords as well as for words. In contrast, the redintegration account suggests that pattern completion mechanisms only operate for words at a late stage, after the degree of phonological degradation has been determined for nonwords, and consequently has more difficulty accounting for the effect of lexical and semantic factors on nonword recall.

Finally, this work found that vowels were recalled more accurately than consonants, in line with several other studies (e.g., Gathercole et al., 2001; Treiman & Danis, 1988), as both migration and identity errors occurred less frequently for these phonemes. The order memory advantage for vowels was equivalent for words and nonwords, suggesting that the greater phonological stability of the words helped to prevent migrations of both vowels and consonants. In contrast, the reduction in identity errors for vowels was smaller for words compared with nonwords. The vowels' acoustic energy may have assisted their identification and maintenance in STM, minimising the number of identity errors involving this phoneme. This effect would have been less critical for words than for nonwords, however, since identification and memory of each word phoneme would have been supported by knowledge about the word itself.

5.3 Experiment 2: Healthy participants tested on pure lists of words and nonwords

The participants in Experiment 1 were tested on mixed lists of words and nonwords. In contrast, in Experiment 2, participants were tested on pure lists of words and nonwords, allowing the effects of mixing words with nonwords to be investigated. In particular, it was possible to determine whether mixing words with nonwords enhances nonword recall as well as having a detrimental impact on word recall. The redintegration account apparently predicts no difference between mixed and pure lists for nonword recall, as redintegration should only improve the recall of words and is thought to occur at a late stage, after the degree of phonological degradation has been determined for nonwords. In contrast, the semantic binding account predicts that the stronger phonological coherence of words should improve the phonological integrity of nonwords by reducing the extent to which nonword phonemes can migrate between list items.

5.3.1 Method

5.3.1.1 Participants

The participants were 20 undergraduates, aged between 18 and 23, who spoke English as a first language and had normal hearing. They were tested individually and took part for course credit.

5.3.1.2 Design and materials

As in Experiment 1, participants were presented with lists of five CVC stimuli. Lexicality, frequency and imageability were retained as within-subjects factors. However, each list was composed purely of words or nonwords, rather than a mixture of the two. The pure word lists were constructed by replacing the nonwords in the mixed lists with real words. Similarly, the pure nonword lists were constructed by replacing the words in the mixed lists with nonwords. Therefore, the participants were tested on the same stimuli as in Experiment 1, in the same serial positions, but in the context of pure rather than mixed lists. The new words, used to replace the nonwords in the mixed lists, had similar frequency counts and imageability ratings as the original Experiment 1 words (see Appendix 11). The new nonwords were constructed from the words they replaced by changing the final consonants (e.g., gun to 'gudge'). Whenever possible, consonants were exchanged between items in order construct the nonwords, although this did not prove feasible for a few common final consonants, given the requirement for every phoneme to be different within a list. All the nonwords were legal and pronounceable. Items were not repeated in the course of the experiment.

There were 60 nonword trials, corresponding to the complete set of lists used in Experiment 1. There were also 40 word trials, corresponding to the lists that contained two and three words in Experiment 1. In the word condition, participants were not tested on lists that originally contained four nonwords and one word because the overlap in items between the two experiments would have been low.

5.3.1.3 Procedure

A female speaker recorded the new words and nonwords individually in a flat intonation. Sound editing software (Cool Edit, Syntrillium) was used to replace either the words or nonwords from the original lists with these new items, in order to produce pure lists of words and nonwords. Presentation was at a rate of one item per second. The pure word and nonword lists were presented in separate blocks and participants were told in advance whether the block would contain words or nonwords. There were four practice trials at the start of each block. The order of the blocks was counterbalanced across participants. The order of trials was the same as for the first twenty participants in Experiment 1. Other methodological details were as described for Experiment 1.

5.3.2 Results

As in Experiment 1, omissions were positioned in response transcripts to minimise the number of errors and on the rare occasions when participants recalled six rather than five items (0.01% of trials for both words and nonwords), the final item was discarded.

5.3.2.1 Recall accuracy

Table 5.2 shows the mean percentage of words and nonwords recalled in the correct serial position in pure and mixed lists. A within-subjects ANOVA was used to compare words with nonwords, with list type (mixed or pure) entered as a between-subjects factor. Only items presented in both the mixed and pure list experiments were included in this analysis, although a separate analysis that examined the recall of the complete set of items yielded a very similar pattern of results. The effect of lexicality was significant ($F(1, 48) = 970.69$, $p < 0.0001$) and there was also a highly significant lexicality by list type interaction ($F(1, 48) = 121.18$, $p < 0.0001$). Planned comparisons revealed that word recall was significantly poorer in mixed compared with pure lists ($t(48) = 4.92$, $p < 0.0001$), suggesting that the presence of nonwords in mixed lists had a detrimental impact on word recall. In addition, nonwords were recalled more accurately in mixed compared with pure lists ($t(48) = 2.13$, $p < 0.05$), suggesting that the greater phonological stability of the words had a knock-on effect on nonword recall.

Table 5.2: *Percentage of words and nonwords in pure and mixed lists recalled by healthy participants (Experiments 1 and 2) and SD patients (Experiment 3)*

		Normal participants				SD patients			
		Pure lists		Mixed lists		Pure lists		Mixed lists	
		M	SD	M	SD	M	SD	M	SD
Words	High frequency, high imageability	87.8	11.5	73.3	11.3	57.6	19.1	44.8	18.6
	High frequency, low imageability	85.8	8.7	68.0	13.9	34.4	18.0	38.4	18.5
	Low frequency, high imageability	74.0	17.0	56.5	14.7	26.4	22.2	28.8	17.5
	Low frequency, low imageability	68.6	18.9	51.5	17.2	26.4	15.6	24.8	10.4
	Mean	79.1	11.6	62.3	11.8	36.2	17.3	34.2	14.9
Nonwords		31.1	11.5	39.4	14.8	-	-	18.3	5.9

Note: only items tested in both mixed and pure lists conditions were included in this analysis. See Tables 5 and 6 for individual patient performance.

A second within-subjects ANOVA was used to examine the effect of frequency and imageability on word recall, with list type (mixed or pure) entered as a between-subjects factor. The main effect of frequency was highly significant ($F(1, 48) = 91.05, p < 0.0001$) and did not interact with list type ($F(1, 48) < 1$), suggesting that frequency had an equivalent effect on the recall of words in mixed and pure word lists. Similarly, the main effect of imageability was significant ($F(1, 48) = 14.37, p < 0.001$) and did not interact with list type ($F(1, 48) < 1$). There were no other significant interactions. It is worth noting that although recall was near ceiling for the pure five-word lists in some conditions, the same pattern of results was obtained when the mixed lists were compared with lists of seven words, tested primarily to provide a comparison with the performance of SD patients (see Experiment 3 below). In this analysis, neither frequency nor imageability interacted with list type ($F(1, 40) < 1$).

By-items analyses, in which lexicality, frequency and imageability were entered as between-items factors and list type (pure or mixed) was entered as a within-items factor, revealed similar patterns of results. The main effect of lexicality was significant ($F(1, 128) = 259.29, p < 0.0001$) and as before, lexicality interacted with list type ($F(1, 128) = 132.15, p < 0.0001$). Bonferroni t tests indicated that word recall was better for pure than mixed lists ($t(99) = 9.46, p < 0.0001$). In contrast, nonword recall was more accurate for mixed than pure lists ($t(179) = 6.44, p < 0.0001$). The main effect of frequency also reached significance ($F(1, 96) = 29.02, p < 0.0001$), although the main effect of imageability did not ($F(1, 96) = 2.22$, n.s.). Again, there were no significant interactions involving frequency or imageability ($F(1, 96) < 1$).

5.3.2.2 Error analysis

The effects of lexicality, frequency and imageability on the number of order and identity errors in pure lists were examined using a series of t tests, as in Experiment 1. In addition, the effect of mixing words with nonwords on these error types was investigated for each factor using a series of repeated-measures ANOVAs, in which list type (i.e., mixed lists from Experiment 1 vs. pure lists from Experiment 2) was entered as a between-subjects variable. Only the pure list items that were also presented in the mixed lists were included in this analysis. As in Experiment 1, lexicality had opposite effects on whole-item identity and order errors in pure lists (see Figures 5.1a and 5.1b). Whole item transpositions occurred more frequently for words than for nonwords ($t(19) = 3.58$, $p < 0.01$), and this lexicality effect did not interact with list type ($F(1, 48) < 1$). Item identity errors were, in contrast, more common for nonwords than for words ($t(19) = 29.31$, $p < 0.0001$), and there was a highly significant lexicality by list type interaction ($F(1, 48) = 1117.83$, $p < 0.0001$). Bonferroni t tests indicated that item identity errors occurred more frequently for words presented in mixed lists compared with pure lists ($t(48) = 10.44$, $p < 0.0001$) and less frequently for nonwords in mixed compared with pure lists ($t(48) = 11.97$, $p < 0.0001$).

As in Experiment 1, a greater number of both order and identity errors occurred for the nonwords compared with the words at the level of individual phonemes (see Figures 5.2a and 5.2b), suggesting that the nonwords were more likely to fragment in verbal STM. Fewer phoneme order errors occurred for words than for nonwords ($t(19) = 7.50$, $p < 0.0001$), and this lexicality effect interacted with list type ($F(1, 48) = 29.50$, $p < 0.0001$). Bonferroni t tests indicated that the lexicality effect was greater for the pure compared with the mixed lists, largely because the number of phoneme migration errors increased substantially for words in mixed lists ($t(48) = 3.58$, $p < 0.001$). In contrast, list type did not affect the number of phoneme migration errors for nonwords ($t(48) < 1$). There were also fewer phoneme identity errors for words than for nonwords ($t(19) = 15.06$, $p < 0.0001$), and this effect interacted with list type ($F(1, 48) = 66.74$, $p < 0.0001$). Bonferroni t tests indicated that the number of phoneme identity errors increased for words when they were mixed with nonwords ($t(48) = 3.18$, $p < 0.01$) and decreased for nonwords when they were mixed with words ($t(48) = 2.73$, $p < 0.01$).

The frequency of the words in the pure lists influenced the numbers of both phoneme order errors ($t(19) = 3.01$, $p < 0.01$) and phoneme identity errors ($t(19) = 5.03$, $p < 0.0001$), providing a replication of the results obtained with mixed lists (see Figures 5.3a and 5.3b). There was no interaction between frequency and list type for either phoneme order or identity errors (both $F(1, 48) < 1$). In addition, the imageability of the words in the pure lists influenced the numbers of phoneme identity errors ($t(19) = 3.34$, $p < 0.01$)

but not phoneme order errors ($t(19) = 1.07$, n.s.), again providing a replication of the results obtained with mixed lists (see Figures 5.4a and 5.4b). There was no interaction between imageability and list type for either phoneme order errors ($F(1, 48) < 1$) or phoneme identity errors ($F(1, 48) = 2.56$, n.s.).

5.3.3 Discussion

In this experiment, participants recalled pure lists of words and nonwords in order to provide a comparison with the mixed lists examined in Experiment 1. As in Experiment 1, recall was better for words compared with nonwords and was superior for words high in frequency/imageability compared with words low in frequency/imageability. These lexical and semantic factors were again found to affect both identity and order errors at the level of individual phonemes. Phoneme identity errors were less frequent for words compared with nonwords, and were less common when word frequency and imageability were high, suggesting that lexical and semantic knowledge constrained the identity of phonemes in STM. In addition, the phonemes of words were less likely to migrate to new list positions than the phonemes of nonwords and similarly, fewer phoneme migration errors occurred for high compared with low frequency words. This result suggests that stable lexical representations constrained the order of phonemes in STM, reducing the incidence of phoneme migration errors, as predicted by the semantic binding hypothesis (Patterson et al., 1994).

An important novel finding from this experiment was that mixing words with nonwords impaired the recall of words and improved the recall of nonwords relative to pure list performance. Both the semantic binding hypothesis and the late-stage redintegration theory can provide an account of the detrimental effect of nonwords on word recall in mixed lists, although their explanations are rather different in nature. One possibility is that the nonword phonemes, which were relatively free to migrate between the list items, damaged the phonological integrity of the words by recombining with their constituent phonemes. This explanation, which is consistent with the semantic binding hypothesis, was put forward to account for the way in which the proportion of words to nonwords affected the degree of phonological disintegration in Experiment 1. Alternatively, in mixed lists, the participants' inability to anticipate which items would be words and nonwords in advance may have prevented the redintegrative mechanism from working normally. Although redintegration is usually thought to be an automatic process (e.g., Hulme et al., 1991), strategic factors may also operate. In pure word lists, participants may deliberately use their knowledge that the target items are real words to constrain their responses. In mixed lists, however, the items are far less predictable as words or nonwords, severely limiting the usefulness of this purposeful lexical reconstruction.

The redintegration theory can therefore account for the smaller number of phoneme order and identity errors in pure word lists compared with mixed lists if it is assumed that redintegration operates more fully for words in pure than mixed lists. It is not clear, however, how the redintegration mechanism discriminates between the degraded traces of words, which need to be reconstructed, and the phonological traces of nonwords, which should not be reconstructed as words, when words and nonwords are presented in unpredictable locations in mixed lists. In addition, the suggestion that redintegrative processes are disrupted as nonwords are added to word lists appears to be incompatible with the Experiment 1 finding that imageability effects were larger when the proportion of words to nonwords was smaller.

Nonwords were also recalled more accurately when they were mixed with words compared with when they were presented in pure nonword lists. This result mirrors the finding from Experiment 1 that the characteristics of words (e.g., word frequency, imageability and the proportion of words to nonwords) affected nonword recall in mixed lists. This finding strengthens the suggestion that the coherence of word items has a knock-on effect on the integrity of the whole phonological trace. If it assumed that redintegration operates on the phonological representations of individual items (e.g., Hulme et al., 1991; Schweickert, 1993), the redintegration theory has difficulty accounting for this finding – the late-stage redintegration of the words in mixed lists should not influence the degree of phonological degradation for nonwords. In contrast, the semantic binding account predicts that the stronger phonological coherence of words should improve the phonological integrity of nonwords by reducing the extent to which nonword phonemes can migrate between list items. However, one potential caveat should be noted. In this experiment, mixing words with nonwords primarily improved nonword recall by reducing phoneme identity errors. The number of phoneme order errors for nonwords was not affected by the presence of words, and yet this is the specific prediction of the semantic binding hypothesis. In contrast, the proportion of words to nonwords in Experiment 1 only affected the number of phoneme order errors for nonwords, and the frequency and imageability characteristics of the words affected both phoneme order and identity errors for nonwords. Of course, the different errors types are not independent of each other, and participants may have failed to show an effect of list type on the number of order errors for nonwords in this experiment because identity errors (98% of which were omissions) prevented the order errors from occurring.

5.4 Experiment 3: Recall of words and nonwords by patients with semantic dementia

The numerous phoneme migration errors made by SD patients in their ISR for words that they no longer fully comprehend suggest that lexical-semantic representations may make a major contribution to the phonological coherence of words in STM. Patterson et al. (1994) noted the apparent similarity between the word recall errors of SD patients and the nonword recall errors of healthy participants and suggested that the phoneme migration errors in the two groups occurred for essentially the same reason, namely, lack of lexical-semantic binding. However, the exact pattern of errors has never been explicitly compared between SD patients and healthy participants. In this experiment, five SD patients were tested on the mixed lists of words and nonwords used in Experiment 1 and the pure lists of words used in Experiment 2, in order to determine whether their errors on pure word lists were qualitatively similar to normal participants' errors on mixed lists of words and nonwords. The mixed list methodology allowed this comparison to be made on the same real word items.

It was also possible to compare the size of the lexicality, frequency and imageability effects for SD patients and healthy participants. As the meanings of lower frequency words generally degrade earlier in the course of SD (Funnell, 1995), it seems likely that the patients' semantic deficits will be more severe for the lower frequency words, leading to the prediction that frequency effects in pure word lists should be larger in the SD patients than in the normal participants. In contrast, lexicality effects should be smaller in the SD patients, particularly when nonwords are compared with low frequency, highly degraded words, as nonword recall should be relatively unaffected by the patients' semantic deficits. It is more difficult to know what to expect for imageability. N. Martin and Saffran (1997) found reduced imageability effects in a group of aphasic patients with semantic impairments, consistent with a reduced semantic contribution to STM. However, Knott et al. (1997) found sizeable effects of imageability in the ISR of a patient with SD, and suggested that the effect may have arisen because of an abnormal difficulty in activating low imageability representations (although the absence of control data in this study makes it difficult to draw firm conclusions).

5.4.1 Method

Five patients were tested in this experiment: SJ, BS, EK, KI and JT. Case description are provided in Section 4.2. The data collection was roughly contiguous with the studies reported in Chapter 4.

The five SD patients were tested on the mixed lists of words and nonwords used in Experiment 1 and the pure words lists used in Experiment 2, allowing a comparison to be made with normal performance. They were not tested on the pure nonword lists from Experiment 2 due to limitations on testing time and the tolerance of the patients. The lists from each condition were interspersed throughout the experiment in order to minimise order effects. Three patients (SJ, KI and JT) were tested using one order of presentation (order A) and two patients (BS and EK) were tested using the reverse order (order B). The presentation order was identical for the mixed and pure word lists. The two sets of lists were presented in different test sessions, separated by a period of at least two weeks.

Further procedural details were as described for Experiments 1 and 2, except that each experiment was divided into two blocks allowing the patients to take a break in the middle of the testing. In addition, while the computer typically controlled presentation of the lists, this was not the case for the earliest patient testing (SJ on the mixed lists). In this initial session, the items were read aloud at a rate of one item per second by the experimenter.

The performance of healthy participants on the pure lists of five words approached ceiling, possibly reducing the size of the frequency and imageability effects. Therefore, twelve additional healthy participants were tested on pure lists of seven words. The participants were undergraduates, aged between 18 and 23, who spoke English as a first language and had normal hearing. They were tested individually and took part for course credit. The lists of seven words were constructed from the original five word lists by adding two additional CVC words in the fourth and fifth serial positions. The new words had similar frequencies and imageability counts to the original words (see Appendix 12). Due to the constraints on the selection of the words and the length of the lists, it did not prove possible to prevent all repetitions of phonemes within the lists, although they were minimised as far as possible (two phonemes or fewer were repeated). Items were not repeated over the course of the experiment. There were 40 trials, corresponding to the 40 word lists tested in Experiment 2. A female speaker recorded the new words individually in a flat intonation. Sound editing software (Cool Edit, Syntrillium) was used to position the new items within the lists of five words used in Experiment 2. The words were again presented at a rate of one item per second. There were four practice trials at the start of the experiment. The order of presentation was the same as for the patients, with equal numbers of participants tested on orders A and B. Other methodological details were as described for Experiment 1.

5.4.2 Results

Omissions were positioned in the response transcripts to minimise the number of errors. BS, EK, KI and JT rarely produced more than five responses for either the mixed lists

(0.8% of trials) or the pure word lists (2.5% of trials) and on the rare occasions when they recalled six rather than five items, the final item was discarded. One patient, SJ, produced six rather than five responses in a third of trials in the mixed lists condition, although she rarely produced more than five responses in the pure word lists condition (2.5% of trials). For this patient, when six responses were produced in the mixed lists condition, the item that minimised the overall number of phoneme errors was discarded.

5.4.2.1 Recall accuracy on pure word lists

Table 5.3 shows the mean percentage of items recalled by the patients and healthy participants on pure lists of five words, as a function of word frequency and imageability. All of the items were included in this analysis, not just those present in the mixed lists of words and nonwords, in order to maximise the amount of data available. The performance of four out of five patients fell more than two standard deviations below the control mean for five-word lists in every condition, suggesting that the ISR of these SD patients was impaired. One patient, BS, who had relatively mild semantic impairments, only showed poor ISR in a single condition.

The number of items recalled by patients and healthy participants on five-word lists were analysed using a within-subjects ANOVA, in which frequency and imageability were entered as within-subjects factors and participant group (patients vs. normal participants) was included as a between-subjects factor. Only effects involving participant group are reported below. The main effect of group was highly significant ($F(1, 23) = 47.14$, $p < 0.0001$), indicating that the patients' recall was poorer than that of normal participants. The interaction between imageability and group approached significance ($F(1, 23) = 3.57$, $p = 0.07$), suggesting that the effect of imageability may have been larger in the SD patients. However, the normal participants' recall was near ceiling, possibly reducing the size of the imageability effect in this group (see below). There was no evidence of a frequency by group interaction ($F(1, 23) < 1$), suggesting that the patients were not more impaired on the lower frequency items overall. The three-way interaction between frequency, imageability and group was significant however ($F(1, 23) = 4.79$, $p < 0.05$).

The frequency by group interaction approached significance for high imageability words ($F(1, 23) = 5.36, p = 0.07$) but not for low imageability words ($F(1, 23) < 1$), suggesting that the patients showed a larger frequency effect than the controls when the words were highly imageable.

Table 5.3: *Percentage of items recalled by SD patients in pure lists of five words (Experiment 3)*

List length	Participant	High frequency		Low frequency	
		High imageability	Low imageability	High imageability	Low imageability
5 items	SJ	28*	26*	16*	12*
	BS	78	64*	54	48
	EK	40*	32*	12*	8*
	KI	60*	38*	34*	28*
	JT	54*	32*	28*	20*
	Patient Mean	52.0	38.4	28.8	23.2
	Patient SD	19.1	14.9	16.6	15.8
5 items	Control Mean	87.7	87.2	75.2	67.0
	Control SD	10.6	8.3	16.8	17.9
7 items	Control Mean	68.1	63.5	56.4	46.7
	Control SD	13.7	14.9	10.5	11.9

* denotes performance more than two standard deviations below the control mean for five-word lists

Note: Table includes data from all items, not just those present in Experiment 1.

A second ANOVA was used to analyse the percentage of words recalled by the patients and healthy participants as a function of frequency and imageability when the controls were tested on seven-item lists to bring their recall off ceiling. These data are shown in Table 5.3. The pattern of results was similar to that described above. Recall was still better for the healthy participants compared with the patients ($F(1, 15) = 10.69, p < 0.01$), even though the healthy participants were presented with longer lists. Although neither frequency nor imageability interacted with group ($F(1, 15) = 2.52$, n.s. and $F(1, 15) < 1$ respectively), there was again a three-way interaction between frequency, imageability and group ($F(1, 15) = 7.01, p < 0.05$). The frequency by group interaction was significant for high imageability words ($F(1, 15) = 4.85, p < 0.05$) but not for low imageability words ($F(1, 15) = 1.98$, n.s.), indicating that the patients showed a larger frequency effect than the controls when the words were highly imageable.

Bonferroni *t* tests were used to further explore the cause of this significant interaction. Percentage recall was lower for the patients compared with the healthy participants in three conditions even though the controls were tested on longer lists: the patients recalled fewer high frequency, low imageability words ($t(15) = 3.16, p < 0.05$), low frequency, high imageability words ($t(15) = 4.18, p < 0.01$) and low frequency, low imageability words ($t(15) = 3.37, p < 0.05$). In contrast, percentage recall did not differ for the patients and controls on high frequency, high imageability words ($t\ (15) = 1.98$, n.s.), indicating that the patients were less impaired in this condition (although Table 5.3 indicates that they were still impaired relative to controls tested on the same length lists). Nonparametric tests (Mann-Whitney U) revealed the same pattern of significant differences.

5.4.2.2 Recall accuracy on mixed lists

Table 5.4 shows the percentage of items recalled by each patient on the mixed lists of words and nonwords as a function of lexicality, frequency, imageability and the proportion of words to nonwords. These data were analysed using a within-subjects ANOVA that incorporated participant group (patients vs. normal participants) as a between-subjects factor. Only effects involving participant group are reported below. There was a main effect of group, indicating that the patients' recall was impaired relative to normal participants ($F(1, 33) = 14.62, p < 0.001$). However, this effect was rather smaller for the mixed lists compared with the pure word lists (partial $\eta^2 = .31$ and $.67$ respectively), consistent with the suggestion that mixing words with nonwords impaired phonological coherence for the normal participants so that their recall was similar to that of SD patients (see below). There was a marginally significant interaction between imageability and group ($F(1, 33) = 4.06, p = 0.05$). Bonferroni t tests indicated that the recall difference between patients and healthy participants was significant for both high imageability items ($t(33) = 4.25, p < 0.001$) and low imageability items ($t(33) = 3.35, p < 0.01$) but was rather smaller when imageability was low, suggesting the patients may not have derived the same degree of benefit from imageability as the controls. Nonparametric tests (Mann-Whitney U) showed the same pattern of results.

Table 5.4: *Percentage of items recalled by SD patients in mixed lists of words and nonwords (Experiment 3)*

		SJ	BS	EK	KI	JT	Patients		Controls	
							M	SD	M	SD
High freq, high imag	W	23*	68	36*	61	21*	41.8	21.5	70.4	14.5
	N	22	35	17	13	22	21.8	8.4	48.2	17.9
High freq, low imag	W	39	53	23*	56	30*	40.2	14.1	63.5	14.3
	N	28	28	9	14	27	21.2	9.2	42.0	16.4
Low freq, high imag	W	29*	43	23*	31	33	32.0	7.3	60.1	15.0
	N	13	17	12	6*	21	13.8	5.6	36.5	14.8
Low freq, low imag	W	16*	36	17	34	26	25.6	9.5	47.2	15.3
	N	9	21	12	16	19	15.6	4.7	31.3	14.2
3:2 words to nonwords	W	25*	55	22*	52	25*	35.7	16.2	66.6	11.9
	N	23	25	10	8	20	17.0	7.8	39.9	15.9
2:3 words to nonwords	W	20*	50	18*	40	33	32.0	13.6	55.9	13.4
	N	18	22	13	17	20	18.0	3.2	39.3	15.4
1:4 words to nonwords	W	35	45	35	45	25*	37.0	8.4	58.3	16.0
	N	14	29	14	13	28	19.3	8.1	39.3	15.4
Grand Mean	W	27*	50	25*	46	28*	34.9	11.9	60.3	12.7
Grand Mean	N	18	25	12	12	23	18.1	5.8	39.5	14.7

* denotes performance more than two standard deviations below the control mean for five-word lists. W = words, N = nonword

5.4.2.3 Comparing recall accuracy for pure and mixed lists

The effect of mixing words with nonwords on each frequency by imageability condition was compared for the patients and healthy participants in a within-subjects ANOVA, in which list type (mixed vs. pure) and participant group (patients vs. normals) were entered as between-subjects factors. This analysis just included the words that were presented in both the pure and mixed lists (see Table 5.2). Only interactions involving participant group are reported below. The three-way interaction between frequency, imageability and participant group reached significance ($F(1, 56) = 5.13$, $p < 0.05$), mirroring the pattern for pure word lists (see above). The interaction between participant group and list type did not reach significance in the data set as a whole ($F(1, 56) = 2.89$, n.s.). However, the four-way interaction between participant group, list type, frequency and imageability approached significance ($F(1, 56) = 3.72$, $p = 0.06$). Consistent with this finding, there was a significant interaction between frequency, imageability and participant group for the pure word lists ($F(1, 23) = 7.38$, $p < 0.05$) but not for the mixed lists ($F(1, 33) < 1$). This pattern suggests that, for pure word lists, the patients' performance was more intact for high frequency and imageability words compared with the other conditions (see the analysis of pure word lists above). In contrast, the patients did not show this lexical-semantic support when the words were mixed with nonwords. Therefore, mixing words with nonwords lessened the involvement of lexical/semantic factors in verbal STM for the patients as well as the healthy participants, but this effect could only be observed for the high frequency, high imageability words which still benefited from lexical/semantic constraints.

5.4.2.4 Error analysis

The effect of frequency and imageability on order and identity errors in the SD patients' recall of pure word lists was examined using the method adopted in Experiments 1 and 2. This analysis only included the original set of items used in Experiment 1. As noted above, it has been proposed that phoneme migration errors occur in the recall of nonwords by normal participants and semantically degraded words by SD patients for the same reason, namely a lack of lexical/semantic binding. If this is the case, SD patients' errors on high and low frequency words (which are expected to be relatively well-known and more semantically degraded respectively) should resemble the errors made on words and nonwords in mixed lists by normal participants.

Four out of five patients made abnormally large numbers of both identity and migration errors at the level of individual phonemes when they were compared with healthy participants tested on five-word lists (see Table 5.5). The one exception was BS, who had relatively mild semantic deficits and more intact verbal STM compared with the other patients. In addition, all five patients made abnormally large numbers of identity errors at the level of whole items. In contrast, item order errors did not occur more frequently for the patients compared with the controls.

Frequency had opposite effects on identity and order errors at the level of whole items for the SD patients, mirroring the influence of lexicality on the recall of normal participants in Experiment 1. Whole-item transpositions occurred more often for high than low frequency words (Wilcoxon test Z (N = 5) = 2.06, $p < 0.05$). In contrast, whole-item identity errors were more common for low compared with high frequency words (Wilcoxon test Z (N = 5) = 2.02, $p < 0.05$).

Identity errors at the level of individual phonemes were also more common for low than high frequency words in the SD patients' recall (Wilcoxon test Z (N = 5) = 2.02, $p < 0.05$). In contrast, there was no overall difference in phoneme order errors for high and low frequency words (Wilcoxon test Z (N = 5) < 1), although there was some suggestion that fewer phoneme migration errors occurred on the high frequency words when only high imageability items were included in the analysis (high frequency mean = 21.4% of items presented, SD = 10.8; low frequency mean = 26.9%, SD = 7.7; Wilcoxon test Z (N = 5) = 1.75, $p = 0.08$). This finding is consistent with the three-way interaction between frequency, imageability and participant group observed in recall accuracy, which showed that the patients' recall was more preserved for high frequency, high imageability words. In addition, the patients made significantly more phoneme migration errors on low than high frequency words when the analysis only included items that were partially recalled (i.e., one or two phonemes correct; high frequency mean = 4.0%, SD = 2.4; low frequency mean = 8.4%, 7.7; Wilcoxon test Z (N = 5) = 2.02, $p < 0.05$). This analysis excluded phoneme order errors that resulted from the migration of whole items, suggesting that the phonemes of high frequency words were more likely to be recalled together in the correct configuration, even if the item was recalled in the wrong place in the sequence.

Table 5.5: *Errors made by SD patients on five-word lists (Experiment 3)*

		SJ	BS	EK	KI	JT	Patients		Controls	
							M	SD	M	SD
High frequency	Phoneme identity	38*	13	28*	18*	22*	23.8	9.7	6.1	4.5
	Phoneme order	31*	10	31*	22*	27*	24.2	8.8	5.9	5.9
	Item identity	65*	25*	52*	42*	52*	47.2	14.9	9.1	5.3
	Item order	13	6	12	8	6	9.0	3.3	4.2	4.5
Low frequency	Phoneme identity	39*	22	29	21	27	27.6	7.2	15.1	9.5
	Phoneme order	30*	4	47*	27*	23	26.2	15.4	9.7	8.4
	Item identity	86*	44	82*	70*	78*	72.0	16.7	24.0	12.9
	Item order	2	0	6	4	0	2.4	2.6	5.0	6.0
High imageability	Phoneme identity	35*	21	30*	18	20	24.8	7.3	9.0	6.7
	Phoneme order	28*	5	40*	21	26*	24.0	12.7	8.3	7.9
	Item identity	69*	31*	70*	46*	58*	54.8	16.5	14.2	8.1
	Item order	11	0	8	4	2	5.0	4.5	5.0	6.4
Low imageability	Phoneme identity	43*	14	27*	21	29*	26.8	10.8	12.1	6.5
	Phoneme order	33*	9	38*	28*	24*	26.4	11.1	7.3	5.9
	Item identity	82*	38*	64*	66*	72*	64.4	16.3	18.9	9.2
	Item order	4	6	10	8	4	6.4	2.6	4.2	4.3

* denotes performance more than two standard deviations above the control mean for pure five-word lists. Errors are expressed as a percentage of items presented.

The effect of imageability on these error types was rather weak for the SD patients. There was some suggestion that item identity errors were less common for high than low imageability words in the SD patients' recall (Wilcoxon test Z (N = 5) = 1.75, $p = 0.08$) but there was no difference between high and low imageability words for item order errors (Wilcoxon test Z (N = 5) < 1), phoneme identity errors (Wilcoxon test Z (N = 5) = 1.22, n.s.) or phoneme order errors (Wilcoxon test Z (N = 5) < 1).

5.4.3 Discussion

Five patients with SD were tested on the pure word lists and mixed lists of words and nonwords used with normal participants in Experiments 1 and 2. The patients' ISR was markedly impaired relative to the performance of healthy participants, particularly for pure word lists, as it was characterised by numerous phonological errors, which occurred much less frequently for the controls. The patients were less strikingly impaired on the mixed lists, presumably because the healthy participants also made abundant phonological errors on the word items under these conditions.

It was expected that the patients' recall would be more impaired for low than high frequency words, as the meanings of low frequency items generally degrade earlier in the course of SD (Funnell, 1995). Lexical-semantic constraints on phonological STM should therefore be diminished for low frequency items, producing larger effects of frequency in SD patients than in controls. The analysis of recall accuracy did provide some support for this prediction, although the frequency effect was modulated by imageability. The patients showed greater effects of frequency than the controls for high but not low imageability words. The patients also showed a larger imageability effect than the healthy participants for high but not low frequency items. In other words, the patients' recall was relatively intact in the high frequency, high imageability condition, but was more substantially impaired in every other condition. These results are largely consistent with a previous study that found strong and interacting effects of frequency and imageability in a patient with SD (Knott et al., 1997).

It is intriguing that semantically impaired patients can show amplified effects of imageability in immediate recall, given that the semantic contribution to STM might be expected to be diminished in these individuals (Martin & Saffran, 1997). However, as discussed by Knott et al. (1997), large effects of imageability may arise in semantically impaired patients because of a difficulty in recruiting semantic representations to support verbal STM for low imageability words. Low imageability items, which are thought to have sparser semantic representations than high frequency words, may be more vulnerable to the effects of semantic degradation (Plaut & Shallice, 1991).

Mixing words with nonwords reduced the impact of lexical/semantic variables on the verbal STM performance of SD patients, relative to pure word lists, in line with the effects observed for normal participants. However, as lexical/semantic factors had a rather weaker influence on the patients' recall of pure word lists when frequency and imageability were low, the effect of mixing words with nonwords was greatest for high frequency and imageability items in this group. For normal participants, the nonword phonemes in mixed lists appeared to impair the phonological coherence of words by recombining with their constituent phonemes. For SD patients, the phonemes of low frequency and imageability words were prone to migrate between items even in pure word lists, and consequently, low frequency/imageability words had little to loose from the presence of nonwords. In contrast, the coherence of high frequency/imageability words was jeopardised by the nonwords in mixed lists in a relatively normal way, and as a consequence, the patients showed reduced effects of imageability in the mixed lists condition, relative to the controls.

The influence of frequency on the patients' recall errors in pure word lists mirrored the effect of lexicality observed for normal participants in Experiments 1 and 2. Frequency, like lexicality for the normal participants, had opposite effects on identity and order errors at level of whole items. The patients made a larger number of item identity errors for low frequency items and more whole-item transpositions for high frequency items. In contrast, both order and identity errors at the level of individual phonemes were more prevalent for low frequency words. Although the overall frequency difference in phoneme migrations was marginal, the frequency effect was more substantial when only partially recalled words were considered (i.e., when phoneme migrations that resulted from whole-item transpositions were excluded), suggesting that the phonemes of high frequency items were more likely to stay together in ISR. Presumably, the larger number of whole-item transpositions occurred for the high frequency words, not because order memory was somehow better for the low frequency items, but because the phonemes of high frequency words were more likely to remain together when they migrated.

This similarity between the effects of frequency for SD patients and lexicality for healthy participants strengthens the suggestion that phonological errors arise for low frequency words and nonwords in the two groups for the same reason, namely a lack of lexical/semantic binding. Highly frequent words (which were likely to be better understood by the patients) were more coherent in STM, suggesting that semantic knowledge constrained memory for phoneme identity and helped to bind the phonemes of less semantically degraded words together, reducing phoneme migration errors.

5.5 General Discussion

Three experiments examined the ISR of both healthy participants and patients with semantic dementia (SD) in order to explore the influence of lexical and semantic factors on the integrity of representations in phonological STM. In the first experiment, healthy participants attempted to recall lists composed of a mixture of words and nonwords. This methodology made it possible to study the effect of lexical/semantic factors on the coherence of the phonological trace for both word and nonword items. In the second experiment, the mixed lists were compared with pure lists of words and nonwords, in order to investigate the impact of the mixed list methodology. In the third experiment, SD patients were tested on these pure and mixed lists, allowing the effect of semantic impairment on immediate recall to be explored. This discussion will first outline the key results of these experiments and will then compare them with the predictions of two accounts of the lexical/semantic contribution to verbal STM: the redintegration theory (Baddeley et al., 1998; Hulme et al., 1991; Hulme et al., 1997; Schweickert, 1993) and the semantic binding hypothesis (Patterson et al., 1994).

There were clear effects of lexical/semantic factors on the recall of normal participants in the mixed and pure lists used in Experiments 1 and 2, as words were recalled more accurately than nonwords and recall was better for more frequent and imageable words compared with their low frequency/imageability counterparts. The role of these factors was examined more closely in an error analysis, in which order and identity errors were considered separately for whole items and individual phonemes. At the level of whole items, lexical/semantic factors had opposite effects on identity and order errors in both Experiments 1 and 2. Identity errors were less common for words compared with nonwords and when frequency and imageability were high, suggesting that lexical/semantic knowledge supported memory for item identity. In contrast, whole-item migration errors were more prevalent for words than nonwords.

At the level of individual phonemes, lexical/semantic factors had similar effects on identity and order errors. In both Experiments 1 and 2, phoneme identity and migration errors occurred less often for words than nonwords and when frequency and imageability were high, suggesting that stable lexical/semantic representations constrained both the identity and ordering of phonemes in STM. Word phonemes were more likely to be recalled together in the correct configuration in STM, whereas the phonemes of nonwords were more likely to fragment. Whole-item order errors were presumably more prevalent for words than nonwords because nonword phonemes rarely migrated as a complete item. Therefore, in line with several other studies (Gathercole et al., 2001; Poirier & Saint-Aubin, 1995, 1996; Saint-Aubin & Poirier, 1999), this work found that lexical/semantic factors primarily benefited memory for identity rather than order information at the level of whole items. However, contrary to the contention of Gathercole et al. (2001), this distinction did not appear to extend to individual phonemes, as lexicality and frequency had a sizeable impact on the occurrence of both phoneme migration and identity errors.

In Experiment 1, frequent phonological errors were observed for the word as well as the nonword items in mixed lists, in contrast with the pure word lists examined in Experiment 2 and the majority of previous studies involving pure word recall (e.g., Henson et al., 1996; Pickering et al., 1998). This result replicates the findings of Knott and Monsell (unpublished study). The nonwords appeared to jeopardise the phonological stability of the words by recombining with their constituent phonemes, causing the word phonemes to migrate. This suggestion was supported by two findings; first, word phonemes were more likely to migrate between the list items when the proportion of nonwords in the mixed lists was greater, and secondly, there were more word-phoneme order errors when the mixed lists were compared with pure word lists.

Interestingly, lexical and semantic variables, namely, frequency, imageability and the proportion of words to nonwords, affected the recall of nonwords as well as words in the mixed lists. All three of these factors influenced the number of phoneme order errors for nonwords, consistent with the suggestion that nonword phonemes had more limited opportunities to migrate when the phonological coherence of the other list items was greater. Similarly, when the mixed and pure lists were compared in Experiment 2, the presence of words in mixed lists was found to boost the recall of nonwords.

In Experiment 3, the SD patients had markedly impaired immediate recall of pure word lists, principally because they made many more phonological errors than healthy participants, suggesting that stable semantic representations play an important role in maintaining the phonological integrity of words in STM. Frequency and imageability effects were exaggerated in the patients' recall, consistent with the view that the meanings of low frequency, abstract words degrade earlier in the course of SD. The high and low frequency words used in this experiment may, therefore, have corresponded to the relatively well-known and more semantically degraded items employed in previous studies (see Chapter 2 for a review). In line with this suggestion, the patients' errors on the high and low frequency words were strikingly similar to those reported for known and degraded items and also resembled the errors made by healthy participants on words and nonwords in Experiments 1 and 2. The patients were poorer at recalling item identity for the low frequency words, as they were more likely to recall the phonemes of these items incorrectly and recombine them with elements from different list items. In contrast, they actually made fewer whole-item order errors for the low frequency items, presumably because the phonemes of low frequency words were less likely to have migrated as a complete item. The patients were rather less impaired on the mixed lists, particularly for low frequency/imageability words, as the healthy participants also made abundant phonological errors on the word items when they were mixed with nonwords.

These results can be condensed into four major findings: 1) lexical and semantic factors, e.g., lexicality, frequency, imageability and semantic degradation in SD, affect recall accuracy, 2) these factors influence both migration and identity errors at the level of individual phonemes, 3) word recall is impaired by the nonwords in mixed lists and 4) nonword recall is enhanced by the words in mixed lists. The following discussion will consider the extent to which the redintegration and semantic binding accounts are consistent with these findings.

As noted in Section 1.3.3, the redintegration theory proposes a two-stage process to account for the role of lexical/semantic factors in verbal STM; a rapidly decaying phonological store and a later reconstructive process, which compares the degraded phonological trace with stable lexical representations in order to reinstate the correct phonological activation (Baddeley et al., 1998; Hulme et al., 1997; Schweickert, 1993; Schweickert, Chen, & Poirier, 1999). According to this approach, lexical/semantic factors influence recall because they affect the efficacy of the redintegration process. The lexicality effect arises because nonwords lack stable phonological-lexical representations, and therefore the redintegration process is largely unable to benefit their recall (Hulme et al., 1991). High frequency words are thought to have more accessible or better-specified lexical representations, which support redintegration particularly effectively (Hulme et al., 1997). In addition, although the reconstructive process is underpinned by phonological-lexical representations, the model can account for semantic effects in immediate recall by assuming that semantic activation contributes to the selection of lexical candidates for reconstruction (Poirier & Saint-Aubin, 1995).

Although the degradation plus redintegration theory clearly predicts that there should be a substantial effect of lexicality in ISR, it is not clear how the redintegration mechanism can discriminate between the degraded traces of words, which need to be reconstructed, and the phonological traces of nonwords, which do not. It might be assumed that metacognitive knowledge enables the system to consistently utilise the redintegration mechanism for pure word lists but not for pure nonword lists. However, when words and nonwords are presented in unpredictable locations in mixed lists, this metacognitive knowledge is likely to be lacking. Consequently, the redintegration theory appears to have some difficulty explaining how effects of lexicality persist in mixed lists. Although it might be possible to argue that redintegration is partially utilised in mixed lists and more fully utilised in pure word lists, this theory is apparently incompatible with the finding that imageability effects become larger as more nonwords are added to the list. In addition, the redintegration framework may have some difficulty accounting for the dramatically impaired recall of SD patients, as it is not clear why the phonological integrity of semantically degraded words should break down in such a major way if phonological-lexical representations are dominant in the redintegration process. This caveat only applies if the long-term associations between phonemic elements, which represent lexical knowledge within the phonological system, remain intact in SD.

The effects of lexical/semantic factors on immediate recall are also easily accommodated by approaches that suggest the language processing system, with its inherent lexical and semantic constraints, underpins verbal STM, and there is no phonological store that can operate independently of lexical and semantic knowledge. As noted in Section 1.3.1, there have been several advocates of this view: Patterson et al.'s (1994) semantic binding hypothesis proposes that verbal STM emerges from interactions between phonological and semantic representations within a parallel distributed processing (PDP) framework. N. Martin and Saffran (1997) suggested that verbal STM results from interactive activation between phonological, lexical and semantic nodes following Dell and

O'Seaghda's (1992) model of speech production and Gathercole and Martin (1996) likened the operation of phonological STM to McClelland and Elman's (1986) TRACE model of speech perception (also see Hulme et al., 1991, for a related suggestion). These approaches all suggest that stable representations of the sounds and meanings of familiar words help to constrain the phonological activation that underpins verbal STM, producing more accurate recall of words than nonwords. These constraints are stronger for more frequent words, giving rise to the frequency effect. Similarly, highly imageable words benefit from stronger semantic constraints. These approaches also readily account for the devastating impact of semantic degradation on the ISR of SD patients, as the interactivity between semantics and phonology can prevent the phonological system from operating normally in the absence of semantic input.

Both theoretical approaches can, therefore, largely account for the impact of lexical and semantic factors on recall accuracy. The second key finding in this chapter, namely that these factors influenced both phoneme migration and identity errors, is also highly consistent with the semantic binding hypothesis. According to this framework, memory for phoneme identity is better for words than for nonwords by virtue of the fact that word phonemes are more strongly associated and will boost each other's activation. In addition, the strong connections between the phonemic elements of familiar words will help to prevent phoneme migration errors by facilitating the binding of phonemes into coherent items.

The redintegration theory also predicts that lexical/semantic factors should influence memory for phoneme identity but does not offer a specific explanation of the impact of these factors on phoneme order errors. Indeed, Gathercole et al. (2001) found that lexicality predominately affected phoneme identity rather than order errors in pure lists, and consequently suggested that "the lexicality effect originates in the redintegration of item information" (p. 1). The redintegration process is purported to restore phonemes in words that have been recalled incorrectly, reducing the number of identity errors. However, as degradation of the phonological trace is assumed to be insensitive to the lexical status of items, an equal number of word and nonword phonemes should migrate.

When phonemes erroneously intrude into words, redintegration can reinstate the correct phoneme, but this process is not expected to correct the identity of word phonemes that intrude into nonwords in mixed lists. Therefore, the predictions of the redintegration theory seem to be at odds with the finding that lexical/semantic factors affect the likelihood of phoneme migration errors. In contrast, the semantic binding hypothesis specifically predicts that the phonemes of words should be less likely to migrate between list items than nonword phonemes.

The third key finding concerned the influence of nonwords on word recall in mixed lists. Both the redintegration theory and interactive models like the semantic binding hypothesis can account for the poorer word recall observed in mixed compared with pure lists, although their explanations are rather different in nature. According to the semantic binding hypothesis, the phonemes of nonwords are not tightly bound together as coherent items, and can therefore act as 'free radicals', recombining with the phonemes of words and damaging their phonological integrity. In order to explain this finding within the redintegration framework, it must be supposed that the reconstructive mechanism cannot operate as effectively for words when they are mixed with nonwords. Although redintegration is usually thought to be an automatic process (e.g., Hulme et al., 1991), strategic factors may also operate, potentially accounting for the difference between pure and mixed lists. In pure word lists, participants may deliberately use their knowledge that the target items are real words to constrain their responses. In mixed lists, however, the items are far less predictable as words or nonwords, severely limiting the usefulness of this purposeful lexical reconstruction. In the current study, the difference in word recall between pure and mixed lists may have arisen either because of the presence of 'unbound' nonword phonemes in the mixed lists, or because of a reduction in deliberate lexical reconstruction. As noted above, however, it seems likely that strategic factors would have been minimised for the mixed lists, suggesting that word recall was poorer in Experiment 1 when the proportion of nonwords was higher because the nonword phonemes directly interfered with the phonological integrity of the words.

The fourth key finding, that the number and lexical/semantic characteristics of the words in mixed lists influenced the recall of nonwords, is also consistent with the semantic binding hypothesis. According to this framework, the coherence of the words in mixed lists will have a major impact on nonword recall, as lexical and semantic constraints will discourage word phonemes from breaking apart in STM, reducing the opportunity for nonword phonemes to migrate. Therefore, this approach anticipates that lexical and semantic constraints relating to specific items should affect the integrity of the entire phonological trace. The redintegration account, on the other hand, suggests that the correct phonology of individual items is reinstated at a late stage. Although this reconstruction process is more effective for highly frequent and/or imageable words, the phonological coherence of nonwords presented with these words should not be enhanced. Consequently, this account has more difficulty explaining the effect of words on nonword recall in mixed lists.

Taken together, these findings support the notion that lexical/semantic factors influence the accuracy of ISR by altering the efficacy of the pattern completion processes that operate for familiar items. Interactive approaches like the semantic binding hypothesis, which suggest that these pattern completion properties are integral to the operation of the phonological system underlying verbal STM, are more consistent with several of the findings reported here than the two-stage degradation plus redintegration theory. The semantic binding hypothesis provides a better account of the impact of lexical and semantic factors on phoneme migration errors, can explain the effect of these variables on nonword recall in mixed lists, readily accounts for the role of semantic as well as phonological-lexical representations in recall and is consistent with the major impact of semantic impairment on ISR observed in SD patients.

6 Lexical and semantic influences on immediate serial recognition

6.1 Introduction

It is informative to consider the effects of lexical and semantic factors on immediate serial recognition performance, as the different theories outlined in Chapter 1 make varying predictions about the time course of the LTM contribution to verbal STM. Some versions of the redintegration hypothesis predict that stable linguistic representations only contribute to verbal STM during the process of recall (e.g., Gathercole, Pickering, Hall, & Peaker, 2001; Schweickert, 1993; Walker & Hulme, 1999). According to this viewpoint, the rapidly decaying phonological trace is impervious to the effects of LTM until the speech output stage, when accurate phonological representations are reinstated for familiar words by comparing the short-term trace with stable lexical-phonological representations. In contrast, the semantic binding hypothesis (Patterson, Graham, & Hodges, 1994) suggests that stable linguistic representations make a contribution to verbal STM throughout immediate recall tasks, by appropriately constraining activation in the phonological system. Lexical and semantic constraints should increase the likelihood of the network settling on the right pattern of phonological activation during encoding and support the maintenance of this activation throughout the task.

This chapter examines the influence of lexical and semantic factors on matching span, a serial recognition task in which two successive lists of items are judged to be the same or different. This task does not require overt recall and so is expected to bypass the redintegration mechanism. Consequently, the redintegration account suggests that lexical influences will be reduced or even abolished in matching span. In line with this prediction, Gathercole et al. (2001) found a much smaller lexicality effect in matching span compared with recall in normal participants. Similarly, Knott et al. (2000) found no difference between words that an SD patient understood relatively well and poorly in matching span, despite finding such a difference in recall. In both of these studies, the matching span task involved detecting transpositions in the order of items (e.g. ABCD became ACBD).

Although the semantic binding hypothesis suggests that lexical and semantic variables contribute to both recall and matching span tasks, it is not incompatible with these findings. The matching span task may have been minimally sensitive to the role of lexical and semantic variables as it required changes in item order to be detected but did not require memory for the items themselves. In contrast, the data presented in Chapter 5 suggest that lexical and semantic factors particularly affect the retention of items rather than their order. Healthy participants recalled nonwords more poorly than words in that study, not because whole items occurred in the wrong positions in the lists, but because their constituent phonemes migrated between list items or were lost altogether. Several previous studies have also found that lexical and semantic variables predominantly affect item rather than order errors (Gathercole et al., 2001; Hulme et al., 1997; Poirier & Saint-Aubin, 1995, 1996; Walker & Hulme, 1999). Similarly, the large numbers of phonological errors made by patients with SD in Chapters 2 – 5 suggest that semantic degradation particularly affects item coherence in STM. Consequently, the matching span tasks employed by Gathercole et al. and Knott et al. may have been relatively unaffected by lexical and semantic factors either because they bypassed a redintegration process operating specifically at recall or because they did not tap the processes supporting the phonological coherence of items in STM. Larger effects of lexical and semantic variables might occur in a recognition task requiring memory for item identity, i.e., requiring maintenance of phoneme order and identity.

In order to explore this possibility, a novel matching span task was devised, which tapped the ability to detect changes in phoneme order that altered item identity. Participants were presented with a list of items like 'bag, rock, sun, hall', followed by a second list like 'bag, *s*ock, *r*un, hall', in which a pair of phonemes had been exchanged between two list items. The influence of lexical and semantic variables in this task was compared with the effect of these factors in a traditional matching span task, in which item order was changed. Both SD patients and healthy participants were examined on these two types of matching span tasks. The SD patients were tested on relatively well-known and semantically degraded words, whereas the normal participants were examined on mixed lists of words and nonwords as a means of assessing the impact of lexicality, word frequency and imageability on matching span performance.

6.2 Matching span in patients with semantic dementia

Matching span tasks were devised for three SD patients, EK, GT and MK, using the relatively well-known and semantically degraded words that were the focus of sections 2.3 and 2.4 in Chapter 2. Case descriptions are provided for these patients in section 2.2. All three patients were found to recall the known words more accurately than the degraded words, although this difference only reached significance for EK when set size was relatively large. The patients made frequent phonological errors in their recall of the degraded words, in contrast with healthy age-matched controls, suggesting that semantic degradation reduced the coherence of these items in STM.

In the following experiments, the patients' ability to detect changes in item order was compared with their ability to detect changes in phoneme order that altered item identity. In both of these conditions, matching span performance was examined for known and degraded words. A greater impact of semantic degradation might be expected in the phoneme order change task, given that the patients' semantic impairments led to frequent phonological errors but did not obviously impinge on memory for the order of items.

6.2.1 Matching span using known and degraded words defined by naming and definitions

6.2.1.1 Method

Three SD patients and nine controls participated in this study. Three controls, matched to each patient on the basis of age and years of education, were tested on each patient's material.

Matching span performance was examined for the known and degraded words defined by naming and definitions, described in section 2.3. For EK and GT, items that were both named and defined correctly were classified as known and items that were neither named nor defined correctly were classified as degraded. For MK, words that were produced correctly in picture descriptions were also included as known. The known and degraded words were matched for frequency on an item-by-item basis and the groups were matched for syllable length. A more detailed description of item selection is provided in Chapter 2.

Two lists of known or degraded words were read aloud in succession, and the participants were asked to judge if the lists were the same or different. The second list could differ from the first in one of two ways: 1) two neighbouring items could be switched in order (e.g., 'piano, rabbit, balloon, bicycle', followed by 'piano, balloon, rabbit, bicycle'), and 2) the onsets of two neighbouring words could be exchanged (e.g., 'piano, rabbit, balloon, bicycle', followed by 'piano, babbit, ralloon, bicycle'). Examples of both types of changes were provided before the start of test.

The known and degraded lists were yoked so that matched pairs of items occurred in the same serial positions. All the lists were presented twice, allowing the item and phoneme order changes to be made on identical lists. Before the lists were constructed, the result of switching onsets between every possible combination of items was established. Some of these changes resulted in real words being produced (e.g., 'cherry, sheep' to 'sherry, cheap'), some resulted in non-words being produced (e.g., 'rabbit, balloon' to 'babbit, ralloon'), and some were impossible because the onsets for the two words were identical (e.g., 'bowl, belt') or because the word began with a vowel sound (e.g., 'elbow'). Impossible changes were discarded, and changes that were impossible in one type of material were avoided for the yoked items in the other. Changes were selected from the remaining possibilities so that an equivalent number resulted in words and non-words for the known and degraded items. An equal number of changes were made at each serial position. The words that changed were placed in the lists first, and the remainder of the lists were constructed by selecting items at random without replacement and re-pooling the items as many times as required.

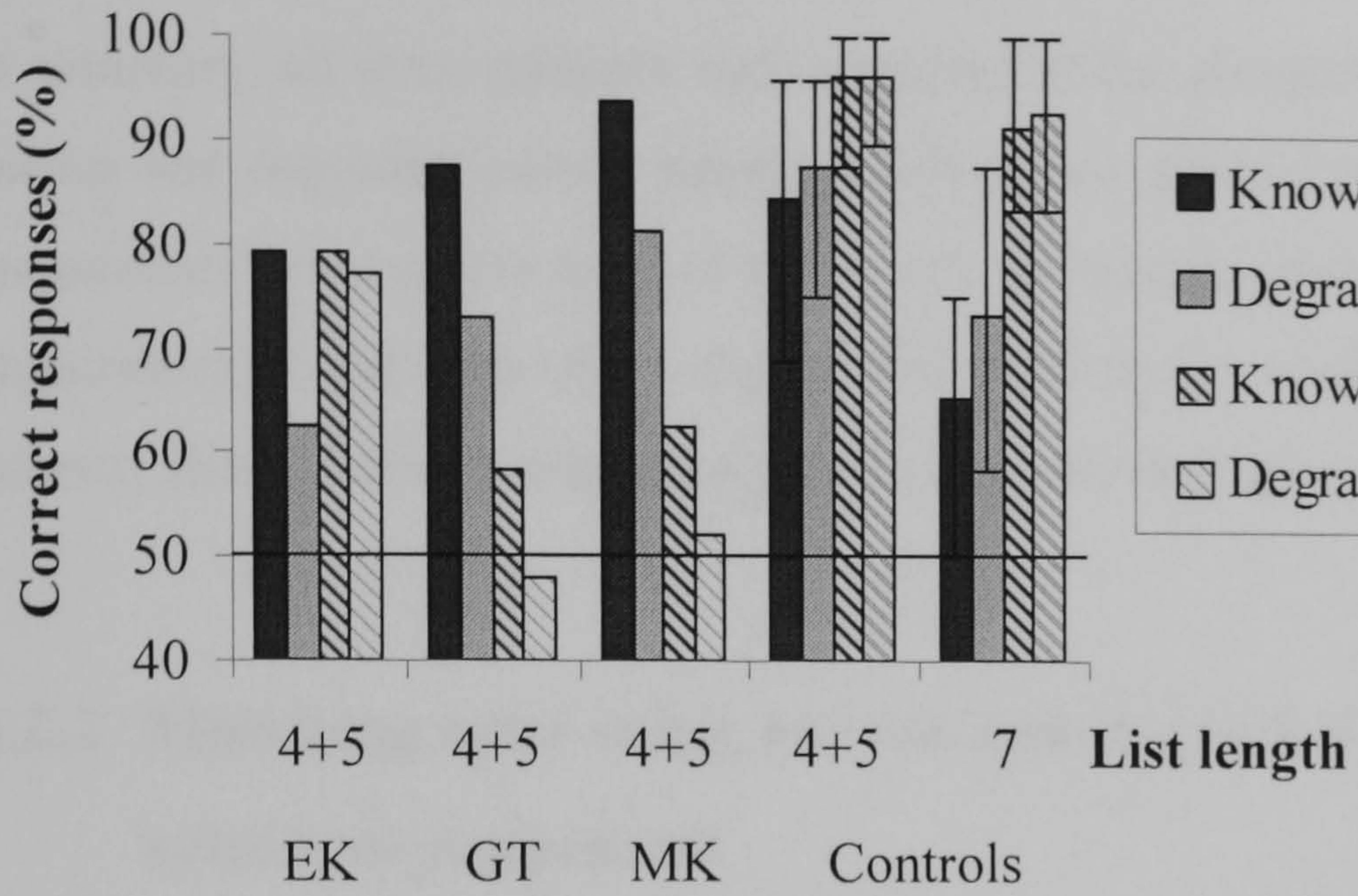
There were four conditions in this experiment, corresponding to the item and phoneme order changes with known and degraded words, each with 24 trials. Changes occurred on half of these trials. The trials were presented in blocks of six trials, arranged using a Latin square design. The patients were tested twice, on lists containing four and five words. The controls were additionally tested on seven item lists. The words were presented at a rate of one word per second, and the two lists were separated by a two second pause. This matching span study was conducted a few months after the recall study reported in Chapter 2.

6.2.1.2 Results

Figure 6.1 shows the percentage of correct responses in the four conditions for the patients and controls, combining the data from four and five item lists. All three patients' detection of item order changes fell within the range of scores obtained by the nine controls on known words. EK and GT's detection of item order changes fell below the normal range for degraded words, although MK's performance was within the normal range. In contrast, all three patients' detection of phoneme order changes fell below the control range, for both known and degraded words.

The control group were better at detecting phoneme than item order changes ($t(8) = 4.65$, $p < 0.01$). In contrast, GT and MK were better at detecting item than phoneme order changes, for both known words (GT: $\chi^2(1) = 8.91$, $p < 0.01$; MK: $\chi^2(1) = 11.98$, $p < 0.001$) and degraded words (GT: $\chi^2(1) = 5.27$, $p < 0.05$; MK: $\chi^2(1) = 7.92$, $p < 0.01$). EK showed a pattern that was more similar to the controls, perhaps because her semantic deficits were milder. Her performance was numerically better for the phoneme order changes, although this advantage did not reach significance for either known words ($\chi^2(1) < 1$) or degraded words ($\chi^2(1) = 1.78$, n.s.).

Figure 6.1: *Matching span for known and degraded words defined by naming and definitions*



Note: chance level = 50%

Error bars show normal range, pooling data from the controls for each patient

Chi-square tests were used to compare matching span performance for known and degraded items. None of the control participants showed a significant difference between known and degraded words; this was the case for both item order change trials (maximum advantage for known words: $\chi^2(1) < 1$ for both 4+5 and 7 item-lists), and phoneme order change trials (maximum advantage for known words: 4+5 item-lists, $\chi^2(1) = 1.90$, n.s.; 7 item-lists, $\chi^2(1) < 1$). In contrast, the known-degraded difference in detecting item order changes approached significance for all three patients in one-tailed tests (EK: $\chi^2(1) = 3.23$, $p = 0.06$; GT: $\chi^2(1) = 3.21$, $p = 0.06$; MK: $\chi^2(1) = 3.43$, $p = 0.06$). The patients did not show a significant known-degraded difference in detecting phoneme order changes, however (EK: $\chi^2(1) < 1$; GT: $\chi^2(1) = 1.05$, n.s.; MK: $\chi^2(1) = 1.06$, n.s.). It is worth noting that floor effects may have reduced the size of the known-degraded difference in phoneme order change trials for GT and MK. Both patients showed a numeric advantage for known over degraded words on four-item lists (GT: 17 vs. 12/24, MK: 17 vs. 13/24) but not on five-item lists (GT: 11 vs. 11/24, MK: 13 vs. 12/24). In addition, the patients as a group did show a significant known-degraded difference in detecting phoneme order changes when the data from this experiment were combined with the results of the following study (see below).

In summary, all three patients were impaired at the phoneme order change task for both known and degraded words, largely because they missed many of the changes (79% of the patients' total errors were of this type). In contrast, none of the patients showed any impairment of the item order change task for known words, although two out of three patients showed abnormally poor performance for degraded words.

6.2.2 Matching span using known and degraded words defined by synonym judgement

6.2.2.1 Method

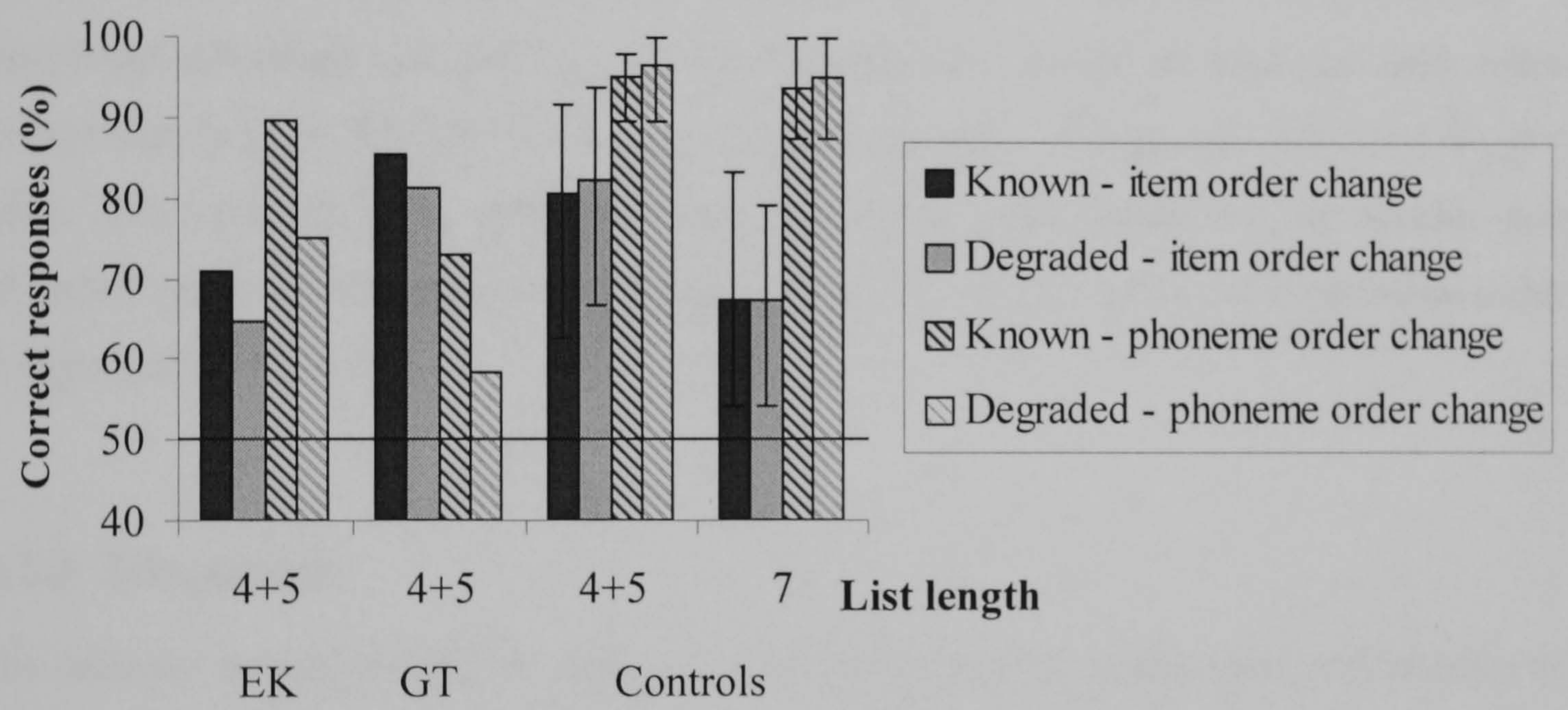
This experiment examined matching span for the known and degraded words defined by synonym judgements described in detail in section 2.4. A synonym judgement task was presented to EK and GT on two separate occasions; items were considered to be 'known' if they were correct on both occasions and 'degraded' if they were incorrect on both occasions. MK was excluded from this experiment as she performed very poorly on the synonym judgement task. The experiment was designed and conducted following the method described in section 6.2.1.1 above. The participants were EK, GT and their six matched controls.

6.2.2.2 Results

Figure 6.2 shows the percentage of correct responses in the four conditions for the patients and controls, combining the data from four and five item lists. EK and GT's detection of item order changes fell within the normal range for known words. Detection of item order changes for the degraded words fell slightly below the normal range for EK but not GT. In contrast, GT and EK were both impaired at detecting phoneme order changes for degraded words. GT's detection of phoneme order changes for known words was below the normal range whereas EK's performance was within the normal range.

The control group were again better at detecting differences in phoneme than item order ($t(5) = 5.36$, $p < 0.01$). EK also showed better detection of phoneme than item order changes for known words ($\chi^2(1) = 4.20$, $p < 0.05$), but not for degraded words ($\chi^2(1) < 1$). As in the previous experiment, GT was better at detecting changes in item than phoneme order for semantically degraded words ($\chi^2(1) = 4.94$, $p < 0.05$), but he did not show this difference for known words ($\chi^2(1) = 1.58$, n.s.).

Figure 6.2: *Matching span for known and degraded words defined by synonym judgement*



Note: chance level = 50%

Error bars show normal range, pooling data from the controls for each patient

None of the control participants showed a significant difference between known and degraded words; this was the case for both item order change trials (maximum advantage for known words: 4+5 item-lists, $\chi^2(1) < 1$; 7 item-lists, $\chi^2(1) = 2.34$, n.s.) and phoneme order change trials (maximum advantage for known words: $\chi^2(1) < 1$ for both 4+5 and 7 item-lists). Neither EK nor GT showed a significant known-degraded difference in detecting item order changes (both $\chi^2(1) < 1$). EK showed a known-degraded difference in phoneme order change trials that approached significance in a one-tailed test ($\chi^2(1) = 3.50$, $p = 0.05$), but GT did not ($\chi^2(1) = 2.26$, n.s.).

In summary, the patients were not substantially impaired at detecting changes in item order, although EK may have been mildly impaired at this task for degraded words. In contrast, both patients showed more substantial deficits in detecting changes in phoneme order for degraded words. GT was also impaired at this task for known words although EK was not. 78% of the patients' errors were failures to detect that a change had occurred.

The three patients as a group showed better performance on the known than the degraded words when the results of the two experiments were combined. They showed a significant advantage for the known over the degraded words on both the item order change lists ($\chi^2(1) = 7.57, p < 0.01$) and the phoneme order change lists ($\chi^2(1) = 5.45, p < 0.05$). The controls as a group showed equivalent performance on the known and degraded items for both item order change lists ($\chi^2(1) = 1.51$, n.s.) and the phoneme order change lists ($\chi^2(1) < 1$).

6.2.3 Discussion

The patients' semantic deficits appeared to influence their matching span performance as well as their ISR (see Chapter 2), suggesting that stable linguistic representations may play a role throughout verbal STM tasks and not just during the process of recall. The patients' semantic impairments particularly affected their detection of phoneme order changes, in line with the view that stable linguistic representations make an important contribution to the coherence of items in verbal STM. GT and MK, who had severe semantic impairments, were markedly impaired at detecting changes in phoneme but not item order. EK, who had milder semantic impairments, showed a more normal pattern of better detection of phoneme than item order changes. Furthermore, the status of the items as known or degraded impacted on the degree to which phoneme order change detection was impaired. In the second experiment, EK was better at detecting phoneme than item order changes for known but not degraded items. In addition, the patients showed a significant known-degraded difference in their detection of phoneme order changes when the data from the two experiments were combined.

The patients' semantic impairments also appeared to influence their detection of item order changes to some extent. EK's detection of item order changes fell below the normal range for degraded but not known words. Moreover, the patients as a group showed a significant known-degraded difference in their detection of item order changes. These findings are consistent with the view that although stable linguistic representations principally affect the coherence of items in STM (i.e., place constraints on phoneme identity and the clustering of phonemes within items), LTM may also contribute to the maintenance of item order. In line with this suggestion, Gathercole et al. (2001) found small but significant effects of lexicality in their matching span task, which tapped memory for item order.

It should be noted, however, that the advantage for known words in matching span did not consistently reach significance for the individual patients. This may have reflected the difficulty of obtaining significant differences in a forced-choice task where guessing alone will permit 50% success. In addition, the difference in performance between the item and phoneme order change conditions for the patients and controls may have been an artefact of the stimulus materials. It was necessary for some of the changes in phoneme order to produce nonwords from real words because of the limited numbers of known and degraded items available for each patient. As a result, it could be argued that this matching span task did not necessarily tap verbal STM; instead, it could have been performed successfully through the detection of nonwords. The control participants may have been considerably more sensitive to changes in lexicality than the patients, given that lexical decision is impaired in SD. What is more, the use of this strategy could have underpinned the SD patients more accurate performance for known compared with degraded words. These concerns are addressed in the following experiment, which examined the influence of lexical and semantic factors on the matching span performance of healthy participants, allowing much tighter control over the stimulus materials.

6.3 Matching span in healthy participants

These studies of matching span in SD patients suggest that clearer influences of stable linguistic representations may occur in serial recognition tasks when participants find it difficult to maintain the phonological coherence of items in verbal STM. In such situations, phoneme identity and migration errors occur more frequently (see Chapters 2 and 5), and as lexical and semantic variables may predominantly act on the occurrence of these types of errors, it seems likely that the role of stable linguistic representations will be more evident. In Chapter 5, we saw that normal participants can be induced to make frequent phonological errors in word recall when they are presented with lists that contain a mixture of words and nonwords. Consequently, in this experiment, the matching span performance of healthy participants was examined on similar mixed lists of words and nonwords, as a way of exploring the influence of lexicality, frequency and imageability on verbal STM prior to recall when the coherence of items was taxed. Participants were examined on both a traditional matching span task, in which items were exchanged in order, and on a task requiring changes in phoneme order, and hence item identity, to be detected. It seems likely that larger effects of stable linguistic representations may emerge on the task tapping memory for phoneme order.

6.3.1 Method

6.3.1.1 Participants

The participants were 72 undergraduates, aged between 18 and 23, who spoke English as a first language and had normal hearing. They were tested individually or in pairs and took part for course credit.

6.3.1.2 Design and materials

Participants were tested on a matching span task, which required the immediate serial recognition of lists of consonant-vowel-consonant (CVC) stimuli. Two five-item lists were presented auditorily and participants decided if the lists had been the same or different. The lists could differ in one of two ways. First, the order of items, but not the items themselves, could differ between the lists. Secondly, the order of the phonemes could be altered, changing the identity of the items. These two types of changes (item order vs. phoneme order) were made on the same lists and were compared as a between-subjects factor.

Each list contained a mixture of two words and three nonwords. Items and phonemes could be exchanged in order between two words, two nonwords or a word and a nonword. Phoneme order changes did not affect the lexical status of the list items. Phoneme exchanges between two words resulted in two new words being produced ('rock, sun' to 'sock, run'), exchanges between two nonwords produced two new nonwords ('leb, hidge' to 'heb, lidge') and exchanges between a word and a nonword produced a word and a nonword ('town, dup' to 'down, tup'). These three levels of lexicality were included as a within-subjects factor.

The influence of word frequency and imageability was also examined. Words were assigned to four high and low frequency by imageability groups on the basis of estimates of written word frequency and imageability taken from the Celex database (Baayen, Piepenbrock, & Rijn, 1993) and the MRC psycholinguistic database (Coltheart, 1981). Mean frequency was 199.6 counts per million for the high frequency (HF) words (range = 60.0 – 1362.1) and 6.9 counts per million for the low frequency (LF) words (range = 0.3 – 19.9). Mean imageability was 585.0 for the high imageability (HI) words (range = 503 – 639) and 413.4 for the low imageability (LI) words (range = 262 – 489). There were no significant frequency differences between the HF HI and HF LI words ($t(118) < 1$) or between the LF HI and LF LI words ($t(118) = 1.23$, n.s.). Similarly, there were no significant imageability differences between the HF HI and LF HI words ($t(118) = 1.58$, n.s.) or between the HF LI and LF LI words ($t(118) = 1.39$, n.s.). The words in each list were taken from a single frequency and imageability group. Phoneme order changes that affected these words resulted in new words of a similar frequency and imageability whenever possible. Frequency and imageability were manipulated for lists that only involved changes to nonwords as well as for word change lists, as it is conceivable that these factors might affect the retention on nonwords in mixed lists of words and nonwords (see Chapter 5). Frequency and imageability were included as within-subjects factors. The items are listed in Appendix 13.

Each participant was tested on 120 lists. There were ten lists in each of the frequency by imageability by lexicality conditions (2 x 2 x 3 conditions), with equal numbers of change and no change trials in each condition. There were two versions of the experiment. In the first version, changes occurred on one set of lists (A) and not on a second set of lists (B). In the second version, the B lists changed and the A lists did not. Half the participants were tested on each version.

The words and nonwords occurred in different serial positions in different lists, in order to prevent the participants from anticipating which items would be words and nonwords in advance. The experiment included five arrangements that maximised the degree to which the words and nonwords were mixed: wnwnn, wnnwn, nwnwn, nwnnw and nnwnw, where 'w' stands for word and 'n' for nonword. The changes were made between nonadjacent items in serial positions 1 and 3, 1 and 4, 2 and 4, 2 and 5, and 3 and 5. There were equal numbers of changes at each of these serial positions for each condition.

The nonwords were constructed from the words by recombining the initial consonants, vowels and final consonants to form new items. Lists were assembled so that vowels were not repeated within a list. Although it was not possible to eliminate all repetitions of consonants, consonants were never repeated in the same syllabic position within a list. Items were not repeated over the course of the experiment.

6.3.1.3 Procedure

The items were recorded individually in a flat intonation by a female speaker and were digitised using a computer. Sound editing software (Cool Edit, Syntrillium) was used to position the items in the lists so that they occurred at a rate of one item per second, with a silent interval of two seconds between the two lists to be compared. Presentation of the lists was controlled using SuperLab software (Cedrus). A red exclamation mark appeared on the computer screen just prior to the start of each trial and remained until the lists had finished playing. It was then replaced by a blue question mark that prompted participants to indicate if the lists had been the same or different. They pressed 'S' on the keyboard if they had detected no change and 'D' to indicate that the lists had been different. The computer recorded their responses. The trials were re-randomised for each participant. The participants were told in advance that the lists would contain both words and nonwords and were given examples of same and different lists. There were six practice trials, on which feedback was given. These trials were presented repeatedly until participants responded correctly. No feedback was given on the experimental trials. The items were presented over headphones.

6.3.2 Results

Table 6.1 shows the number of same and different lists that were correctly recognised as a function of lexicality, word frequency, imageability and change type (whether item or phoneme order was changed). An analysis of variance (ANOVA) was performed on the number of correct trials in each condition to examine the influence of these factors. There were significant main effects of lexicality ($F(2, 140) = 3.30$, $p < 0.05$) and word frequency ($F(1, 70) = 9.04$, $p < 0.01$), but not imageability ($F(1, 70) < 1$). Participants showed better recognition of lists containing high than low frequency words and were better at detecting changes that involving words than nonwords. There was a highly significant between-subjects effect of change type ($F(1, 70) = 55.36$, $p < 0.0001$), indicating that participants were better able to detect changes in item order than phoneme order. This difference may have resulted from the fact that in the former case, three phonemes, including a vowel, were repositioned in the list. In contrast, the phoneme order change task involved the transposition of a single consonant.

There was a significant three-way interaction between imageability, lexicality and change type ($F(2, 140) = 5.54$, $p < 0.01$). Bonferroni t tests indicated that participants showed a significant effect of lexicality for high but not low imageability words in their recognition of phoneme but not item order changes. In the phoneme order change data, participants were more accurate at detecting changes involving two words compared with a word and a nonword for high imageability lists ($t(35) = 4.66$, $p < 0.001$) but not low imageability lists ($t(35) < 1$). They showed no difference between changes involving two nonwords and a word and a nonword, for either high or low imageability lists ($t(35) < 1$). In the item order change data, participants showed no significant effects of lexicality for either high or low imageability words. They showed no difference between changes involving two words and a word and a nonword ($t(35) < 1$ for both high and low imageability lists). They also showed no difference between changes involving two nonwords and a word and a nonword (high imageability: $t(35) = 2.16$, n.s.; low imageability: $t(35) = 1.58$, n.s.). This interaction is evident in Table 6.1. Although the effect of lexicality was small in the data set as a whole (there was only a 1% difference in recognition accuracy for item order changes involving two words and two nonwords), it was rather larger for high imageability words in the phoneme order change condition (4-13%). The two-way interaction between imageability and lexicality contained within this three-way interaction also reached significance ($F(2, 140) = 3.55$, $p < 0.05$). None of the other interactions terms reached or approached significance.

An additional by-items analysis replicated many of these findings. Again, there were significant main effects of change type ($F(1, 108) = 239.94$, $p < 0.0001$) and word frequency ($F(1, 108) = 5.06$, $p < 0.05$), although the main effect of lexicality did not reach significance ($F(2, 108) = 1.28$, n.s.). There was no main effect of imageability (1, 108) < 1). As before, the three-way interaction between imageability, lexicality and change type was significant ($F(2, 108) = 4.61$, $p < 0.05$), suggesting that imageability had a larger influence on the detection of word changes involving exchanges in phoneme order (and consequently, changes in item identity) than exchanges in item order. No other interactions reached or approached significance.

6.3.2.1 Serial position effects

Three separate ANOVAs were used to examine the recognition of changes occurring at each serial position as a function of frequency, imageability and lexicality. Change type was included as a between-subjects factor. These analyses grouped changes involving the initial item (exchanges between items 1 and 3, and 1 and 4), changes involving the final item (exchanges between items 2 and 5, and 3 and 5) and changes involving items in the middle of the list (exchanges between items 2 and 4).

Table 6.2 shows recognition as a function of frequency and serial position. There was no main effect of serial position (by subjects: $F(2, 140) < 1$; by items: $F(2, 114) < 1$). There was a significant interaction between frequency and serial position by subjects ($F(2, 140) = 4.65$, $p < 0.05$) although this effect did not reach significance by items ($F(2, 114) = 1.67$, n.s.). By-subjects Bonferroni t tests indicated that there was no significant frequency effect for changes at the beginning of the lists ($t(71) < 1$). In contrast, there was a significant frequency effect for changes in the middle of the lists ($t(71) = 4.05$, $p < 0.001$) and the frequency effect for changes at the end of the lists approached significance ($t(71) = 2.34$, $p = 0.07$). The three-way interaction between frequency, serial position and change type approached significance (by subjects: $F(2, 140) = 2.75$, $p = 0.07$; by items: $F(2, 144) = 2.54$, $p = 0.08$), apparently because the two-way interaction was stronger for phoneme order change lists (see Table 6.2).

Table 6.1: *Mean percentage of lists correctly recognised as a function of lexicality, frequency, imageability and list type (same or different)*

		Phoneme order changes								Item order changes							
		Word-word		Word-nonword		Nonword-nonword		Mean		Word-word		Word-nonword		Nonword-nonword		Mean	
Word type	List type	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
High frequency, high imageability	Same	70.0	25.5	65.0	23.6	74.4	23.7	71.1	24.8	81.7	21.0	77.8	19.0	83.3	18.8	83.7	18.7
	Different	78.9	23.4	71.1	22.6	67.2	28.7			87.2	19.2	87.2	15.2	85.0	18.1		
High frequency, low imageability	Same	66.7	22.4	75.6	24.0	71.1	21.1	70.6	23.4	83.9	23.3	81.1	16.5	80.6	18.8	83.5	18.3
	Different	75.0	21.6	65.6	24.2	69.4	26.4			87.2	15.2	85.6	17.0	82.8	18.0		
Low frequency, high imageability	Same	73.3	25.3	60.6	23.2	60.0	24.4	67.2	24.3	76.1	20.7	75.0	22.1	80.0	17.9	81.0	19.2
	Different	73.3	19.7	66.7	23.9	69.4	26.8			84.4	16.6	82.2	20.2	88.3	14.6		
Low frequency, low imageability	Same	65.0	26.3	67.2	24.4	65.0	22.6	66.8	24.6	83.3	16.2	82.2	14.9	71.7	22.6	81.4	19.1
	Different	66.7	27.0	66.1	23.8	70.6	24.6			82.2	22.3	84.4	18.0	84.4	17.3		
Mean		71.1	24.2	67.2	23.8	68.4	24.9			83.3	19.6	81.9	18.2	82.0	18.7		

Table 6.2: *Mean percentage of lists correctly recognised as a function of frequency and serial position*

		Item order		Phoneme order	
	Serial position	M	SD	M	SD
High frequency	1 and 3	79.86	12.50	65.28	18.53
	1 and 4	85.88	10.32	71.53	13.57
	2 and 4	86.81	11.85	72.69	12.38
	2 and 5	78.01	15.06	72.22	13.66
	3 and 5	87.50	13.44	72.45	11.41
Low frequency	1 and 3	79.17	16.96	67.59	15.53
	1 and 4	81.48	11.97	70.83	15.49
	2 and 4	82.41	12.08	63.19	12.81
	2 and 5	80.56	13.06	65.28	13.58
	3 and 5	82.41	14.33	68.06	13.44

Table 6.3 shows recognition of change lists as a function of lexicality and serial position. No-change trials, in which the lists to be compared were the same, were excluded from the analysis, given that every list contained a mixture of both words and nonwords. There was a significant interaction between lexicality and serial position (by subjects: $F(4, 280) = 23.26, p < 0.0001$; by items: $F(4, 111) = 11.34, p < 0.0001$). By subjects Bonferroni t tests indicated that there was a significant advantage for word-word changes, compared with word-nonword changes at the end of the lists ($t(71) = 6.04$, $p < 0.0001$). This lexicality effect did not occur at the beginning of the lists ($t(71) = 1.88$, n.s.) and was reversed for changes in the middle of the lists ($t(71) = -4.74, p < 0.0001$). There were no significant differences between word-nonword and nonword-nonword changes at any serial position ($t(71) < 1.67$, n.s.). The three-way interaction between lexicality, serial position and change type did not reach significance (by subjects: $F(4, 280) < 1$; by items: $F(4, 111) < 1$). In addition, there was no significant interaction between imageability and serial position (by subjects: $F(2, 140) < 1$; by items: $F(2, 114) < 1$).

Table 6.3: *Mean percentage of lists correctly recognised as a function of lexicality and serial position*

		Item order		Phoneme order	
	Serial position	M	SD	M	SD
Words-words	1 and 3	66.67	19.71	57.78	23.80
	1 and 4	69.44	17.56	64.44	19.78
	2 and 4	56.11	23.33	44.44	18.58
	2 and 5	74.44	12.29	64.44	15.20
	3 and 5	74.44	12.29	62.78	17.34
Word-nonwords	1 and 3	71.11	13.04	52.78	24.91
	1 and 4	64.44	17.31	57.78	23.80
	2 and 4	73.89	15.73	57.22	24.91
	2 and 5	55.00	23.11	44.44	19.78
	3 and 5	75.00	12.07	57.22	19.80
Nonwords-nonwords	1 and 3	52.78	23.00	51.11	25.94
	1 and 4	74.44	11.32	57.78	21.26
	2 and 4	76.11	11.53	58.89	17.85
	2 and 5	64.44	18.58	53.33	24.38
	3 and 5	72.78	13.65	55.56	23.96

6.3.3 Discussion

This experiment examined matching span in healthy participants when words were mixed up with nonwords. Under these circumstances, the normal participants showed a similar pattern of performance to the SD patients, as they were poorer at detecting changes in phoneme than whole-item order. They showed the opposite pattern to the healthy controls in the neuropsychological studies reported above. These controls were tested on pure word lists and were considerably better at detecting changes in phoneme than item order. It seems likely that at least some of the healthy participants' difficulty with phoneme order changes in this experiment stemmed from the use of mixed lists of words and nonwords. This methodology appears to reduce the stability of phonological representations of words in STM (see Chapter 5), making it difficult to detect single phoneme transpositions. The item order changes would have been relatively easy to detect, even given this phonological instability, as a larger number of phonemes switched position. However, this suggestion should be addressed experimentally in future work, as other methodological differences between the studies may have contributed to the difference in the detection of item and phoneme order changes (i.e., the age of the participants, closed vs. open sets and the lexical status of the items resulting from phoneme order changes).

Small but significant effects of frequency and lexicality were found in the experiment as a whole, suggesting that lexical and semantic variables contributed to this verbal STM task even though it did not involve overt recall. Larger lexicality effects occurred for the more highly imageable words, and more importantly, the effects of lexicality and imageability were larger in the phoneme order change task compared with the item order change task. In other words, the standard matching span task employed by Gathercole et al. (2001) and Knott et al. (2000) proved to be relatively insensitive to the effects of lexical and semantic variables. The novel matching span task that required the detection of phoneme order changes was more sensitive to these variables. This finding suggests that stable linguistic factors improve the coherence of familiar words in STM *prior* to recall (i.e., reduce the extent to which phonemes from different items become confused).

It should be noted, however, that the lexicality, frequency and imageability effects were small in this experiment, even in the phoneme order change condition, when compared with the influence of these variables in ISR for mixed lists of words and nonwords. In Chapter 5, the size of the lexicality, frequency and imageability effects in ISR were approximately 19%, 14% and 12% respectively for mixed lists containing two words and three nonwords. In the phoneme change condition of this experiment, these effects were reduced to an average of 4%, 4% and 0% (this figure for lexicality contrasts word-word changes with nonword-nonword changes and only incorporates data from trials in which a change occurred; the figures for frequency and imageability include both same and different trials). Therefore, the results of this experiment are entirely consistent with the hypothesis that an additional redintegrative process comes into play during overt recall. However, this study does not provide conclusive evidence for late-stage lexical reconstruction. Differences in effect size between serial recall and recognition are difficult to interpret, especially when the items used in the experiments are not identical and when there are complex interactions between factors that are not carefully controlled across the studies. In addition, matching span might be generally insensitive to the effects of experimental manipulations because each trial provides relatively little data, floor and ceiling effects are encountered rapidly and because guessing alone permits 50% success. On the other hand, Gathercole et al. (2001) did find comparable effect sizes for phonological similarity in the two paradigms.

6.4 General Discussion

This chapter has examined the effect of lexical and semantic factors on matching span performance in both SD patients and healthy undergraduates. The traditional matching span task, which primarily taps item order memory, was compared with a novel matching span task requiring changes in phoneme order (and consequently item identity) to be detected. The results were largely consistent across the neuropsychological and normal experimental studies. In situations in which it was hard for the participants to maintain the coherence of words in STM (either because the words were semantically degraded, in the case of SD patients, or because the words were mixed up with nonwords, in the case of healthy undergraduates), changes in item order were detected more readily than changes in phoneme order. As healthy undergraduates make frequent phonological errors, rather than item order errors, when recalling mixed lists of words and nonwords (see Chapter 5) and SD patients show this pattern when recalling semantically degraded words (see Chapters 2, 3 and 4), the participants were poor at detecting changes in matching span that mirrored their recall errors. The SD patients were impaired at detecting phoneme order changes for 'known' as well as 'degraded' words, probably because the semantic degradation underlying the known-degraded distinction is continuous, and the patients' comprehension of the 'known' words was partially compromised.

In contrast with the SD patients and undergraduates tested on mixed lists of words and nonwords, the normal controls in the neuropsychological studies were significantly better at detecting changes in phoneme than item order. It seems likely that this was because the coherence of the words in STM was not compromised by the presence of nonwords. Phonological coherence may have also been boosted by the relatively small set sizes and longer word lengths used in these experiments. The controls would have become familiar with all of the items as they were repeated over the course of the experiment, reducing the likelihood of phoneme migration and identity errors (see Chapter 2 for further discussion). In addition, the multisyllabic words presented to the controls may have been less susceptible to phonological errors than the CVC items used with the undergraduates, as they had fewer potentially confusing phonological neighbours (but see Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002, for evidence suggesting that dense phonological neighbourhoods improve rather than impair recall). In fact, the controls made very few phonological errors in their recall of these words, probably due to a combination of these factors, and in contrast item order errors occurred much more frequently (see sections 2.3 and 2.4). Therefore, the controls were also poorest at detecting matching span changes that mirrored their recall errors.

Lexicality, frequency and imageability were found to affect the matching span performance of healthy undergraduates, particularly in the phoneme order change task that was expected to tap the phonological coherence of items in STM. Similarly, semantic degradation was found to affect the matching span performance of SD patients. Given that the matching span tasks did not require overt recall, these findings are most consistent with the view that lexical and semantic variables contribute to the phonological coherence of words *throughout* STM. The results are less consistent with the notion of a phonological store that is immune from the effects of lexical and semantic variables until a late redintegration process comes into play during recall. The results also suggest that the matching span tasks used by Gathercole et al. (2001) and Knott et al. (2000) may have been relatively insensitive to the effects of lexical and semantic factors, as they required changes in item order to be detected. In contrast, lexical and semantic variables appear to make a particular contribution to the coherence of items in STM (i.e., memory for phoneme identity and the clustering of phonemes within items). Therefore, task sensitivity, and not the absence of overt recall, may have caused the very limited effects of lexical and semantic variables observed in previous studies.

The studies presented in this chapter are consistent with the possibility that lexical and semantic factors play a greater role in recall compared with recognition, although it should be noted that it may be problematic to compare effect sizes between experiments with very different methodologies (see section 6.3.3). If such a difference does emerge, this might suggest that lexical and semantic factors play an enlarged role during the recall process, although would not challenge the suggestion that these factors also play a role throughout STM. The semantic binding hypothesis (Patterson et al., 1994) is not incompatible with possibility that stable linguistic representations make a greater contribution to recall than recognition tasks, as the requirement to actively produce the target items may make recall particularly sensitive to the effects of lexical and semantic support.

Finally, although these results are consistent with the view that stable linguistic factors particularly affect the phonological coherence of words in STM, the SD patients showed a known-degraded difference on the item order matching span task. Similarly, healthy undergraduates showed some effect of lexical and semantic variables on the detection of item order changes, suggesting that stable linguistic factors may contribute to the maintenance of item order, as well as to the coherence of individual items. Given that lexical and semantic factors appear to affect the extent to which phonemes are successfully clustered as an item (i.e., contribute to memory for phoneme order at least over short distances), these variables may have a knock-on impact on item order memory, as the order of phonemes also represents the order of items. In addition, even though the traditional matching span task apparently does not explicitly require participants to remember item information, they may still use item information to perform the test (see

Neath, 1997, for a similar argument). However, these findings are not incompatible with the view that other mechanisms, more immune to the effects of stable linguistic factors, also contribute to the maintenance of item order, accounting for the smaller effects of lexical and semantic variables on tasks that predominantly tap item order memory (this possibility is discussed at greater length in Chapter 7).

7 Discussion and conclusions

7.1 Introduction

The work presented in this thesis has focused on the relationship between stable linguistic representations and verbal STM. This section reviews the key findings from each chapter and discusses some common themes running through them. The results will be compared with the predictions of two accounts of the relationship between LTM and verbal STM: the redintegration perspective (e.g., Baddeley, Gathercole, & Papagno, 1998; Hulme, Maughan, & Brown, 1991; Hulme et al., 1997; Schweickert, 1993) and interactive theories (e.g., Martin & Saffran, 1997; Patterson, Graham, & Hodges, 1994). The primary difference between these viewpoints concerns the extent to which verbal STM is seen as an independent cognitive system.

As noted in Chapter 1, the interactive approach views verbal STM as resulting from activation within the levels of representation that underpin linguistic processing (e.g., phonological, lexical, semantic). These levels interact so that activity in the phonological nodes is constrained by lexical and semantic knowledge. There have been several advocates of this view; for example, Patterson et al.'s (1994) semantic binding hypothesis proposes that verbal STM emerges from interactions between phonological and semantic representations within a parallel distributed processing (PDP) framework, and N. Martin and Saffran (1997) suggested that verbal STM results from interactive activation between phonological, lexical and semantic nodes following Dell and O'Seaghda's (1992) model of speech production.

In contrast, the redintegration theory posits separate short and long-term phonological stores – verbal STM is therefore seen as a more autonomous system (e.g., Baddeley et al., 1998). According to this perspective, stable phonological representations are used to reinstate the rapidly decaying phonological trace during the process of recall (Gathercole, Pickering, Hall, & Peaker, 2001; Hulme et al., 1991; Hulme et al., 1997; Schweickert, 1993). Although this reconstructive process is underpinned by phonological-lexical representations, the model can account for semantic effects in ISR by assuming that semantic activation contributes to the selection of lexical candidates for reconstruction (Poirier & Saint-Aubin, 1995).

Although it is convenient to make comparisons between these two types of account, it should be noted that there is in fact a diversity of 'redintegration' and 'interactive' theories and their features overlap. The discussion below therefore focuses on specific predictions and the extent to which they are compatible with the results obtained in Chapters 2 to 6.

7.2 Review of key findings

7.2.1 Immediate serial recall of relatively well-known and semantically degraded words

Chapters 2, 3 and 4 examined ISR for relatively well-understood and more semantically degraded words in patients with semantic dementia. In Chapter 2, known and degraded words were selected using naming, definition and synonym judgement tests. In line with previous studies, the patients showed poorer recall of the degraded words and made more phonological errors on these items. Chapter 3 considered ISR for number and non-number words. In every SD patient, the recall of single-digit number words was normal whereas the recall of non-number words was impaired relative to controls. This difference extended to lower frequency multi-digit numbers, and remained even when frequency, imageability, word length, set size and size of semantic category were matched across the two sets of words. Additional assessments suggested that comprehension was considerably better for the number than non-number words. The patients' category specific advantage for numbers in ISR may therefore have had a semantic locus – it appeared to be equivalent to the known/degraded recall difference observed in Chapter 2.

This association between semantic impairment and the emergence of phonological errors in ISR suggests that semantics may play a major role in maintaining phonological integrity in normal recall (Patterson et al., 1994). Interactive accounts of the LTM contribution to verbal STM can readily account for the harmful impact of semantic impairment on ISR, as semantic representations are thought to constrain and maintain activation at the phonological level. This interpretation remains controversial, however. A number of studies have failed to find recall differences between relatively well-understood and semantically degraded words in SD patients (Funnell, 1996; Lambon Ralph & Howard, 2000; McCarthy & Warrington, 1987, 2001; Warrington, 1975), leading some authors to argue that verbal STM can operate without the involvement of semantics. By this argument, additional phonological or lexical impairments underlie the known-degraded differences observed in some patients (Knott, Patterson, & Hodges, 1997; McCarthy & Warrington, 2001). Phonological deficits might cause verbal STM to be more reliant on input from semantics, and consequently, patients with both phonological and semantic difficulties should only have accurate ISR for words that are still understood relatively well. It is important to note, however, that all SD patients make frequent phonological errors in ISR, even if they do not show the expected known-degraded difference in recall accuracy. Moreover, inconsistency in the size of known-degraded recall difference could be a consequence of discrepancies in methodology. Consequently, Chapter 2 investigated the effect of various methodological factors on the size of the known-degraded recall difference and Chapter 4 examined the evidence for phonological-lexical deficits independent of the primary semantic deficit in SD.

7.2.2 Methodological factors affecting the size of the known-degraded recall difference

In Chapter 2, the number of known and degraded words in the lists (set size) was found to influence the size of the known-degraded recall difference. Patient EK showed a significant known-degraded difference when set size was large but not when it was small. The frequency of phonological errors fell when the same items were presented repeatedly, suggesting that increased familiarity with the degraded words improved their phonological coherence. Many of the prior studies that found no recall difference between known and degraded words also used small set sizes, suggesting that some of the discrepant results in the literature might be attributable to this factor.

Two patients, EK and GT, were included in these investigations and set size did not affect them in the same way. Smaller set sizes particularly improved EK's recall of the degraded words. In contrast, for GT, decreased set size either had little effect or enhanced his recall of the known words more than the degraded words. Therefore, GT's recall was better for the known words at every set size. These findings point to an effect of semantic knowledge on the degree to which repeatedly presenting words boosts their coherence in STM. Small set sizes might most strongly benefit words that are partially semantically degraded but not completely forgotten. The richness of the two patients definitions suggested that EK's degraded words fell into this category, whereas the meanings of GT's degraded words were more substantially lost. These results are consistent with a study showing that SD patients can relearn the phonological forms of words that they still partly know but are much less able to relearn words that have completely impoverished semantic representations (Snowden & Neary, 2002). However, the relationship between semantic knowledge and phonological learning in SD is an interesting topic that remains relatively unexplored.

Both the redintegration and interactive theories are able to account for the effect of set size on ISR. Roodenrys and Quinlan (2000) proposed that the number of lexical candidates in the redintegrative process is reduced when set size is small, increasing the likelihood that the phonological trace will be reconstructed accurately. Similarly, from an interactive perspective, the repeated presentation of items from limited sets might be expected to increase lexical/semantic activation for those particular items, boosting the constraints on phonology. These theories may also be able to account for the differential effect of set size on known and degraded words. Set size might be expected to have little impact on words with very severely degraded lexical/semantic representations if activation of these same lexical representations underpins the set size effect in healthy participants (Roodenrys & Quinlan, 2000). Set size might also be expected to have little impact on the recall of very well known words, because these items will be adequately supported by their intact semantic representations. Partially degraded words might derive the most support from small set sizes if repeated presentation of these words is able to boost any residual lexical-semantic activation that still plays a role in maintaining phonological coherence in STM. Roodenrys and Quinlan (2000) found that set size only affected the recall of lower frequency items in normal participants, presumably because higher frequency items were adequately supported by their more accessible lexical representations. This result may be analogous to the effect of set size on the recall of well-known and partially degraded words in patients with SD.

Chapter 2 also compared two rather different methods of selecting words as known and degraded: naming and definitions, and synonym judgement tests. Although recall accuracy showed the same pattern across these methods, a known-degraded difference in phonological errors emerged for the naming and definitions words but not for the synonym judgement words. As the semantic degradation underlying the known-degraded distinction varies continuously, the point of cut-off between 'known' and 'degraded' items may have differed between these methods. The 'known' words defined by synonym judgements may have been more degraded than those defined by naming and definitions, accounting for the occurrence of phonological errors on both known and degraded words. This finding suggests that the methods used to select words as known and degraded can influence the size of the recall difference between them.

The results presented in Chapter 2 suggest that most SD patients will show a recall advantage for known words when methodological conditions are favourable (in particular, when set size is large). However, it is not clear from this work whether methodological factors can account for all of the previous failures to find known-degraded recall differences in SD: patient MNA (McCarthy & Warrington, 2001), for example, had excellent ISR for words that she did not understand, despite the large set size used in this study. Therefore, Chapter 4 took a closer look at the possibility that additional phonological or lexical impairments are responsible for the known-degraded recall differences observed in some studies.

7.2.3 Evidence for intact phonology in semantic dementia

SD patients are generally considered to have intact phonology, because they rarely, if ever, make phonological errors in spontaneous speech, picture naming or single-word repetition (Knott et al., 1997). This issue remains relatively unexplored, however, despite its theoretical importance. Chapter 4 examined the performance of six SD patients on a range of phonological processing tasks, e.g., rhyme judgement and production, minimal pairs and phoneme segmentation. The patients with the mildest semantic impairments were unimpaired, whereas the patients with more severe semantic deficits showed some weaknesses. Although this pattern is consistent with the view that the atrophy underlying SD impacts on phonological as well as semantic representations, an alternative possibility, given that the tasks required the maintenance and manipulation of phonological representations, is that the patients' semantic deficits impaired their performance. The tasks would not have been immune from the impact of semantic degradation if semantics does provide a major source of constraint on phonological activation.

Similarly, the patients largely showed normal effects of phonological similarity and word length in ISR. These effects are considered to be hallmarks of normal phonological coding and articulatory rehearsal in verbal STM respectively (e.g., Vallar & Papagno, 2002), although explanations of the word length effect remain controversial (see Chapter 4). There was some suggestion of a reduction in the size of the phonological similarity effect in the most severely impaired patients, consistent with the view that the phonological system is compromised in the latter stages of SD. However, several recent studies have found that phonological similarity interacts with lexicality in normal ISR (Gathercole et al., 2001; Lian, Karlsen, & Winsvold, 2001). Consequently, semantic impairments might be expected to produce a reduction in the size of the phonological similarity effect in the absence of additional phonological deficits.

It might be expected that SD patients' ISR impairments should not extend to nonwords, as the semantic system plays less of a role in maintaining these items. The majority of the patients in Chapter 4 did show normal recall of both single multisyllabic nonwords and strings of monosyllabic nonwords. Moreover, all the patients showed a reduction of the normal lexicality effect. There was again some suggestion of a mild weakness in nonword recall in the patients with the most severe semantic impairments. Importantly, however, nonwords that were phonologically similar to well-known words were recalled more accurately than nonwords that were similar to more semantically degraded words. This ISR difference points to a semantic contribution to nonword recall that could account for nonword recall deficits in SD. Therefore, there was no convincing evidence of an independent phonological impairment in SD across a range of phonological processing and ISR tasks in Chapter 4.

One finding in Chapter 2 also supported the suggestion that SD patients have largely intact phonological STM. List length did not affect the size of the known/degraded recall difference or the incidence of phonological errors. In contrast, the number of non-phonological errors (predominantly omissions) rose steadily as list length was increased. Therefore, phonological coherence was affected primarily by the type of material to be retained, i.e., the status of the words as known or degraded, and not by the amount of material. This finding is consistent with the suggestion that although semantic degradation impairs ISR performance in SD, the underlying phonological STM mechanism is intact. The patients were able to retain phonological representations of a relatively normal number of items in STM, although the coherence of these representations was weakened for degraded words, allowing phonological elements to migrate between the list items. This result is highly consistent with the prediction that semantics helps to bind the phonemes of words together in verbal STM (Patterson et al., 1994) but does not specifically support the view that interactive-activation between phonology and semantics plays an important role in maintaining the level of phonological activation above a threshold for recall (Martin & Saffran, 1997).

All of the patients examined in Chapter 4 showed a significant known/degraded recall difference, whether or not they performed normally on phonological tasks. This finding supports the view that semantics makes a contribution to the coherence of items in verbal STM in the absence of any additional phonological impairment. Evidence was also obtained to suggest that the known/degraded recall difference is underpinned by a central semantic deficit rather than a separate impairment of lexical representations. The ISR advantage for known over degraded items extended to a non-verbal delayed picture-copying task. The patients were able to reproduce more of the correct features when they made delayed copies of drawings that represented their known items, compared with their degraded items. As the same items were impaired in both ISR and delayed copying, it seems likely that degradation of a unitary semantic system caused the difficulties in both tasks (see Rogers et al., in press, for a similar argument). This reasoning applies equally to ISR for words and nonwords; therefore, the patients' deficits in nonword recall apparently resulted from their marked semantic difficulties and not from any additional independent impairment of lexical representations.

This evidence of a semantic contribution to nonword recall is more consistent with interactive than redintegration accounts of the relationship between LTM and verbal STM. In Chapter 4, it was suggested that there were two mechanisms underlying the known/degraded difference for nonwords. The first of these apparently fits with both theories but is insufficient to account for all of the data. The second mechanism is less consistent with the redintegration framework. First, nonwords that closely resemble known words could be recalled more accurately than those derived from degraded words because specific semantic representations of the known item are activated by the nonword's phonology. Within the framework of interactive theories, this semantic activation helps to constrain the phonological segments of the nonword that overlap with the known word. Similarly, activation of phonologically similar lexical representations could allow segments of the nonword to be redintegrated. This semantic contribution to nonword recall is expected to be greatest for multisyllabic items, as long nonwords that are highly phonologically related to particular known words and few other words should produce strong and coherent activation of specific semantic representations relating to those individual words. EK did show a larger ISR difference between 'known' and 'degraded' nonwords for longer items and normal participants might also be expected to exhibit semantic effects for longer nonwords. In contrast, it is less clear how semantic activation could helpfully constrain the phonological trace of short nonwords that have many phonological neighbours, because semantic and phonological representations are uncorrelated.

The finding that several of the more severely semantically impaired patients did show significant known-degraded differences on the shorter nonwords suggests a second mechanism might be operating. The PDP framework (Patterson et al., 1994; Plaut & Kello, 1999) allows for an effect of semantic impairment on nonword recall, even for short items, as according to this approach, stable phonological representations are acquired in the presence of semantics. Consequently, the phonological space can change when the input from semantics is damaged, jeopardising the coherence of semantically degraded items in ISR tasks (see Chapter 4). The redintegration theory has greater difficulty accounting for this semantic contribution to the recall of nonwords with many phonological neighbours.

7.2.4 Lexical and semantic influences on phonological coherence in normal participants

Chapter 5 also addressed the influence of lexical and semantic knowledge on the phonological coherence of items in STM but focused on the recall of healthy participants rather than SD patients. Normal participants rarely make phonological errors in word recall. Instead, whole-item order errors constitute the majority of ISR errors in many studies (e.g., Henson, Norris, Page, & Baddeley, 1996; Pickering, Gathercole, & Peaker, 1998). However, healthy subjects make frequent phoneme migration errors in their recall of nonword lists (Treiman & Danis, 1988). These errors appear to be similar to those made by SD patients on word lists (Patterson et al., 1994). In Chapter 5, when normal participants were presented with lists composed of a mixture of words and nonwords, they made frequent phonological errors in ISR for both types of item. This technique, therefore, encouraged healthy participants to make errors on word items like those displayed by SD patients. Words in mixed lists were recalled more poorly than those in pure lists because their phonemes were more likely to migrate between the list items and phoneme identity errors were more frequent when nonwords were present. Word phonemes were also more likely to migrate when the proportion of nonwords in the mixed lists was greater.

The mixed list methodology made it possible to examine the influence of lexical and semantic factors, namely, lexicality, frequency and imageability, on phonological integrity in healthy subjects. At the level of whole items, lexical/semantic factors had opposite effects on identity and order errors. Identity errors were less common for words compared with nonwords and when frequency and imageability were high, suggesting that lexical/semantic knowledge supported memory for item identity. In contrast, whole-item migration errors were more prevalent for words than nonwords. At the level of individual phonemes, lexical/semantic factors had similar effects on identity and order errors. Phoneme identity and migration errors occurred less often for words than nonwords and when frequency and imageability were high, suggesting that stable lexical/semantic representations constrained both the identity and ordering of phonemes in STM. Word phonemes were more likely to be recalled together in the correct configuration in STM, whereas the phonemes of nonwords were more likely to fragment. Whole-item order errors were more prevalent for words than nonwords presumably because nonword phonemes rarely migrated as a complete item.

Interestingly, lexical and semantic variables, e.g., frequency, imageability and the proportion of words to nonwords, affected the recall of nonwords as well as words in the mixed lists. Phoneme order and identity errors were less frequent for nonwords that were presented in lists containing high frequency/imageability words compared with lists containing low frequency/imageability words. Similarly, nonwords were recalled more accurately when they were presented in mixed lists of words and nonwords compared with pure nonword lists, although it is possible that this effect resulted from a strategic difference brought about by the discrepancy in difficulty between the experiments. Such a strategic difference is less likely to account for the finding that the proportion of words to nonwords affected nonword recall within the mixed-lists experiment, however, as the easy and more difficult lists (containing more and fewer words respectively) were presented in a random order and participants were unable to anticipate which items would be words and nonwords in advance.

Several of the findings from Chapter 5 appear to be more consistent with interactive models than with aspects of the redintegration viewpoint, although many of the results can be accounted for by both theoretical approaches. Both perspectives would predict an influence of lexical/semantic factors on phoneme identity errors. Interactive models predict that phonemes that occur together in words become associated in the phonological system, reinforcing the correct pattern of activation and helping to prevent other phonemes from reaching the threshold for recall. Likewise, lexical reconstruction would allow missing or incorrectly represented phonemes to be reinstated (e.g., 'c*l*ocodile' could be recalled correctly as 'crocodile'). The semantic binding hypothesis (Patterson et al., 1994) also explicitly predicts an effect of lexical/semantic factors on phoneme order errors. The stable associations between the phonemes making up a word encourage them to emerge together in recall, reducing phoneme migrations between list items. In contrast, the redintegration viewpoint does not appear to offer a specific explanation of the influence of lexical/semantic factors on phoneme order errors. Equal numbers of word and nonword phonemes might be expected to migrate if it is assumed that degradation of the phonological trace is insensitive to the lexical status of items and that instead lexicality effects result from a subsequent reconstructive process.

Both the redintegration theory and interactive models like the semantic binding hypothesis can account for the poorer word recall observed in mixed compared with pure lists, although their explanations are rather different in nature. According to the semantic binding hypothesis, the phonemes of nonwords are not tightly bound together as coherent items, and can therefore act as 'free radicals', recombining with the phonemes of words and damaging their phonological integrity. In order to explain this finding within the redintegration framework, it must be supposed that the reconstructive mechanism cannot operate as effectively for words when they are mixed with nonwords. Although redintegration is usually thought to be an automatic process (e.g., Hulme et al., 1991), strategic factors may also operate, potentially accounting for the difference between pure and mixed lists. In pure word lists, participants may deliberately use their knowledge that the target items are real words to constrain their responses. In mixed lists, however, the items are far less predictable as words or nonwords, severely limiting the usefulness of this purposeful lexical reconstruction.

The semantic binding hypothesis can also account for the finding that the number and lexical/semantic characteristics of the words in the mixed lists influenced the recall of the nonwords they were presented with. According to this framework, the coherence of the words in mixed lists will have a major impact on nonword recall, as lexical and semantic constraints will discourage word phonemes from breaking apart in STM, reducing the opportunity for nonword phonemes to migrate. The redintegration account, on the other hand, may have greater difficulty in explaining the influence of lexical/semantic factors on nonword recall in mixed lists, if it is assumed that degradation of the phonological trace is insensitive to the lexical status of items and that lexicality effects come about through a discrete late stage reconstruction process which only operates for words. The results of Chapter 5 suggest that pattern completion processes that operate for familiar words influence the integrity of the whole phonological trace and this finding places useful constraints on models of verbal STM.

Chapter 5 also examined the performance of SD patients on the same pure word lists and mixed lists of words and nonwords that the healthy participants were tested on. The SD patients had markedly impaired immediate recall of pure word lists, principally because they made many more phonological errors than healthy participants. Frequency and imageability effects were exaggerated in the patients' recall, consistent with the view that the meanings of low frequency words degrade earlier in the course of SD (Funnell, 1995). The high and low frequency words used in this experiment may, therefore, have corresponded to known and more semantically degraded items. In line with this suggestion, the patients' errors on the high and low frequency words were strikingly similar to those made by healthy participants on words and nonwords. The patients were poorer at recalling item identity for the low frequency words, as they were more likely to recall the phonemes of these items incorrectly and recombine them with elements from different list items. In contrast, they actually made fewer whole-item order errors for the low frequency items, presumably because the phonemes of low frequency words were less likely to have migrated as a complete item. The patients were rather less impaired on the mixed lists, particularly for low frequency/imageability words, as the healthy participants also made abundant phonological errors on the word items when they were mixed with nonwords.

7.2.5 Lexical and semantic effects on matching span

As noted in Chapter 6, the redintegration and interactive accounts make different predictions about the effect of lexical/semantic variables on immediate serial recognition tasks like matching span. Some versions of the redintegration hypothesis (e.g., Gathercole, Pickering, Hall, & Peaker, 2001; Schweickert, 1993; Walker & Hulme, 1999) suggest that the rapidly decaying phonological trace is impervious to the effects of LTM until the speech output process, when accurate phonological representations are reinstated for familiar words by comparing the short-term trace with stable lexical-phonological representations. According to this account, matching span, which bypasses the recall process, should be unaffected by lexical/semantic variables. In contrast, interactive theories suggest that stable linguistic representations make a contribution to verbal STM throughout immediate recall tasks, by appropriately constraining activation in the phonological system. Consequently, this viewpoint predicts that lexical/semantic influences should be evident in matching span.

Previous studies of matching span have found little influence of lexicality in normal participants (Gathercole et al., 2001) and no effect of semantic degradation in SD patients (Knott, Patterson, & Hodges, 2000). However, these studies used the standard matching span procedure, which involves detecting changes in item order. As lexicality, frequency, imageability and semantic impairment primarily affected identity and not order errors at the level of whole items in Chapter 5, typical matching span tasks may underestimate the involvement of lexical/semantic factors in verbal STM. Therefore, in Chapter 6, the standard matching span task was compared with a novel task in which participants attempted to detect when the onsets of items had been exchanged, altering item identity. SD patients were tested on known and semantically degraded words, whereas normal participants were examined on mixed lists of words and nonwords as a means of assessing the impact of lexicality, word frequency and imageability on matching span performance.

In situations in which it was hard for the participants to maintain the coherence of words in ISR (either because the words were semantically degraded, in the case of SD patients, or because the words were mixed up with nonwords, in the case of healthy undergraduates), changes in item order were detected more readily than changes in phoneme order. In contrast, the healthy participants who acted as controls for the SD patients were much better at detecting changes in phoneme than item order. These participants were presented with the same words repeatedly in pure lists, benefiting the coherence of items in STM.

Lexicality, frequency and imageability affected the matching span performance of healthy undergraduates, particularly in the phoneme order change task that was expected to tap the phonological coherence of items in STM. Similarly, the degree of semantic degradation affected matching span in SD patients. These results are consistent with the notion that stable linguistic factors affect the phonological coherence of words in STM. The effect of these factors was not restricted to the phoneme order change task, however. Both the SD patients and healthy participants showed an effect of lexical/semantic variables on the detection of item order changes, suggesting that stable linguistic factors may contribute to the maintenance of item order, as well as to the coherence of individual items (see below for further discussion).

7.2.6 Item vs. order memory

The results presented here largely support the view that, at the level of whole items, the main impact of lexical/semantic factors is on identity rather than order errors. In Chapters 2 to 5, both SD patients and healthy participants made fewer phonological (i.e., identity) errors on items that received greater lexical/semantic support. In contrast, whole item transpositions were either unaffected by lexical/semantic factors or were actually more common for words with stronger lexical/semantic support. These findings concur with those of a number of previous studies suggesting that lexical/semantic variables predominantly affect item not order errors (Gathercole et al., 2001; Poirier & Saint-Aubin, 1995, 1996; Saint-Aubin & Poirier, 1999; Turner, Henry, & Smith, 2000). However, lexical/semantic influences were observed in a matching span task that required the detection of changes in item order (Chapter 6), suggesting that memory for item order is not entirely immune from these effects. Of course, order and identity errors are not independent of each other. It may have been difficult for participants to detect item transpositions in the matching span task if the identity of items was very poorly represented.

At the level of individual phonemes, lexical/semantic factors appeared to have an impact on both identity and order errors in ISR. These results are compatible with the view that stable linguistic knowledge helps to prevent the migration of phonemes between list items and encourages the correct phonemes to be recalled within items. Phoneme misorderings and identity errors both contribute to the greater incidence of phonological errors for items that lack lexical/semantic support.

Lexical/semantic factors may encourage the elements of familiar words to emerge together in ISR preventing phoneme migration errors because the constituents of words co-occur in speech production and comprehension, and therefore become associated. It is less clear how lexical/semantic factors could constrain the order of whole unrelated word items in ISR because item order is novel and therefore poorly supported by long-term associations. In fact, when word phonemes do move to the wrong serial position, lexical/semantic binding will tend to make the entire word migrate. Of course, this does not hold for sentence repetition tasks, in which word order may be supported by long-term syntactic and conceptual representations (Lombardi & Potter, 1992; Potter & Lombardi, 1990, 1998).

The crucial issue for the interactive and redintegration accounts is whether there is a phonological STM system, independent of stable phonological-lexical representations, that can maintain the serial order of items (i.e., a 'phonological loop'). Such a system would be incompatible with the basic tenet of the interactive perspective – namely, that verbal STM emerges from the representations that underpin language processing. The phonological loop hypothesis is certainly consistent with the finding that lexical/semantic factors predominantly affect item and not order errors at the level of whole items. In addition, most computational models of serial recall (e.g., Burgess & Hitch, 1992; Page & Norris, 1998; see section 1.3.4) incorporate separate mechanisms for recalling the order of items and the items themselves. However, it is important to note that although the differential impact of lexical/semantic factors on item and order errors suggests that mechanisms independent of linguistic representations may contribute to serial recall, this is not evidence for a phonological loop system per se. Instead, a myriad of cognitive resources could play a crucial role in maintaining the order of unrelated words in verbal STM. Some of these functions may be specific to language; for example, syntactic mechanisms might play a role in maintaining word order. Others may be domain-general mechanisms for representing the serial order of events (see Brown, Hulme, & Preece, 2000) or frontal-attentional mechanisms for maintaining activation of relevant posterior representations (Ruchkin, Grafman, Cameron, & Berndt, in press). These mechanisms are expected to be involved in a range of cognitive processes, not just verbal STM, and crucially do not entail a short-term copy of information activated in long-term memory; therefore, this proposal is not incompatible with the interaction perspective. Future research is needed to clarify the extent to which these serial order mechanisms interact with lexical/semantic knowledge.

7.3 Conclusions

7.3.1 The nature of verbal short-term memory: support for an interactive perspective

The work presented in this thesis suggests that stable linguistic representations play an important role in the maintenance of item integrity in verbal STM. Knowledge of the sounds and meanings of familiar words supports accurate ISR by promoting recall of the correct phonemes and encouraging the segments of well-known items to emerge together. The results are most consistent with interactive accounts that view pattern completion properties in verbal STM as constraints within the systems that underlie linguistic processing. They are less consistent with the notion that the phonological trace is initially impervious to lexical/semantic influences and is refreshed at a late stage by separate long-term phonological representations.

Several findings across the chapters specifically support aspects of the interactive account:

- The ISR of SD patients is markedly impaired by semantic impairment, even in the absence of additional phonological deficits, suggesting that interactions between semantics and phonology are crucial to the normal functioning of verbal STM.
- Evidence from SD patients and normal participants suggests that lexical/semantic constraints prevent phoneme migrations as well as phoneme identity errors. Interactive accounts, in particular the semantic binding hypothesis (Patterson et al., 1994), explicitly predict that these factors will encourage the phonemes of words to be recalled together as a coherent item.
- In ISR, lexical/semantic representations of familiar words affect not only the integrity of the items that they specifically relate to but also the coherence of the entire phonological trace.

- A task that is thought to be predominately phonological, namely nonword recall, showed an influence of semantic degradation in SD patients, consistent with the notion that the phonological system does not function normally in the absence of intact semantic input.
- Lexical/semantic effects are not restricted to ISR but also influence matching span performance. This result seems incompatible with the claim that lexical/semantic factors only have an influence during overt recall.

7.3.2 Directions for future research

The above findings are theoretically informative because they illustrate phenomena that successful models of verbal STM should be able to account for. However, it would also be highly worthwhile to compare experimental data sets like those presented in Chapter 5 with implemented interactive and late-stage redintegrative models. Such an approach would highlight the extent to which each account is consistent with this empirical data set and could potentially identity problematic aspects of each of theory. It is clearly difficult to know exactly what the predictions of the two accounts are without implemented models because the descriptive theories themselves are rather poorly specified. It seems unlikely, for example, that the interactive accounts of Patterson et al. (1994) and Martin and Saffran (1997) could perform ISR tasks without modification as they are based on models of single word production and comprehension.

The technique of presenting mixed lists of words and nonwords as a means of inducing healthy participants to make more frequent phoneme migration errors in ISR certainly has potential for future research. This methodology permits an investigation of the impact of lexical and semantic variables on phonological integrity in STM – although healthy participants make frequent phoneme migration errors in nonword recall, lexical/semantic variables cannot be easily manipulated in such studies. The experiments presented in Chapter 5 investigated the impact of lexicality, frequency and imageability, but other factors, including set size and phonological neighbourhood size, could be gainfully explored. It would also be interesting to explore interactions between word length/phonological similarity and list type (pure words vs. mixed lists of words and nonwords). Long words might be more resilient than short words to the presence of nonwords in a list because they have substantially fewer phonological neighbours and this might make their phonemes less likely to recombine. In this way, recall might actually be superior for long compared with short words. Phonological similarity, on the other hand, is expected to increase the number of phoneme and whole item migration errors for words (see Gathercole et al., 2001), because the dissimilar portions of phonologically similar items are more likely to be exchanged. As the mixed list methodology also encourages more frequent phoneme order errors, the detrimental effects of phonological similarity may be lessened. (This hypothesis is related to the suggestion, made in Chapter 4, that SD patients may not show normal effects of phonological similarity even in the absence of phonological deficits). It would also be interesting to establish whether phoneme migration errors are more likely to occur using this technique when the stimuli are selected so that recombinations of phonemes produce real words rather than nonwords. Finally, the role of strategic redintegrative processes could be explored by comparing ISR for lists composed of an unpredictable jumble of words and nonwords with the recall of lists containing words and nonwords in alternate positions.

The matching span results presented in Chapter 6 also point to some directions for future research. The normal participants tested on mixed lists of words and nonwords were better at detecting changes in item order than phoneme order, presumably because a larger number of phonemes moved in this condition. The patients' controls, on the other hand, were tested on pure word lists with a small set of repeating items and were much better at detecting changes in phoneme order. Future studies could investigate factors that influence the relative difficulty of these two conditions, including list type (mixed words and nonwords vs. pure lists of words) and set size. It seems likely that the difficulty of detecting phoneme order changes in matching span is influenced by the phonological coherence of items in STM.

References

Allport, A. (1984). Auditory-verbal short-term memory and conduction aphasia. In D. Bouwhuis (Ed.), *Attention and performance X: Control of language processes* (pp. 313-325). Hillsdale, NJ: Lawrence Erlbaum Associates.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX Lexical Database [CD-ROM]*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417-423.

Baddeley, A., Vallar, G., & Wilson, B. (1987). Sentence comprehension and phonological memory: Some neuropsychological evidence. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 509-529). London: Lawrence Erlbaum Associates.

Baddeley, A., & Wilson, B. A. (2002). Prose recall and amnesia: Implications for the structure of working memory. *Neuropsychologia, 40*, 1737-1743.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18*, 362-365.

Baddeley, A. D. (1972). Retrieval rules and semantic coding in short-term memory. *Psychological Bulletin, 78*, 379-385.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D., & Ecob, J. R. (1970). Simultaneous acoustic and semantic coding in short-term memory. *Nature, 227*, 288-289.

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158-173.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

Baddeley, A. D., & Levy, B. A. (1971). Semantic coding and short-term memory. *Journal of Experimental Psychology, 89*, 132-136.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology, 36*, 233-252.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575-589.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language, 44*, 568-591.

Barnard, P. (1985). Interacting cognitive subsystems: A psycholinguistic approach to short-term memory. In A. Ellis (Ed.), *Progress in the psychology of language, Vol. 2* (pp. 197-258). London: Erlbaum.

Belleville, S., Caza, N., & Peretz, I. (in press). A neuropsychological argument for a processing view of memory. *Journal of Memory and Language*.

Bird, H., Lambon Ralph, M. A., Seidenberg, M., McClelland, J. L., & Patterson, K. (2003). Deficits in phonology and past-tense morphology: What's the connection? *Journal of Memory and Language, 48*, 502-526.

Bourassa, D. C., & Besner, D. (1994). Beyond the articulatory loop - a semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin & Review, 1*, 122-125.

Bozeat, S., Lambon Ralph, M. A., Graham, K., Patterson, K., Wilkin, H., Rowland, J., et al. (2003). A duck with four legs: Investigating the structure of conceptual knowledge using picture drawing in semantic dementia. *Cognitive Neuropsychology, 20*, 27-47.

Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia, 38*, 1207-1215.

Bozeat, S., Lambon Ralph, M. A., Patterson, K., & Hodges, J. R. (2002). The influence of personal familiarity and context on object use in semantic dementia. *Neurocase, 8*, 127-134.

Bransford, J. D., & Johnson, M. A. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717-726.

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology, 26*, 467-482.

Brooks, J. O., III, & Watkins, M. J. (1990). Further evidence of the intricacy of memory span. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 1134-1141.

Brown, G. D. A., Hulme, C., & Preece, T. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*, 127-181.

Burgess, N., & Hitch, G. J. (1992). Towards a network model of the articulatory loop. *Journal of Memory and Language, 31*, 429-460.

Burgess, N., & Hitch, G. J. (1996). A connectionist model of STM for serial order. In S. Gathercole (Ed.), *Models of short-term memory* (pp. 51-72). Hove, UK: Psychology Press.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*, 551-581.

Butterworth, B., Campbell, R., & Howard, D. (1986). The uses of short-term memory: A case study. *Quarterly Journal of Experimental Psychology, 38A*, 705-737.

Butterworth, B., Cappelletti, M., & Kopelman, M. (2001). Category specificity in reading and writing: The case of number words. *Nature Neuroscience, 4*, 784-786.

Cantor, J., & Engle, R. W. (1993). Working-memory capacity as long-term activation: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19*, 1101-1114.

Cappelletti, M., Butterworth, B., & Kopelman, M. (2001). Spared numerical abilities in a case of semantic dementia. *Neuropsychologia, 39*, 1224-1239.

Cappelletti, M., Kopelman, M., & Butterworth, B. (2002). Why semantic dementia drives you to the dogs (but not to the horses): A theoretical account. *Cognitive Neuropsychology, 19*, 483-503.

Caza, N., Belleville, S., & Gilbert, B. (2002). How loss of meaning with preservation of phonological word form affects immediate serial recall performance: A linguistic account. *Neurocase, 8*, 225-273.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.

Cipolotti, L., & Butterworth, B. (1995). Towards a multiroute model of number processing: Impaired transcoding with impaired calculation skills. *Journal of Experimental Psychology: General, 124*, 375-390.

Cipolotti, L., Butterworth, B., & Denes, G. (1991). A specific deficit for numbers in a case of dense acalculia. *Brain, 114*, 2619-2637.

Cipolotti, L., & Warrington, E. K. (1995). Semantic memory and reading abilities: A case report. *Journal of the International Neuropsychological Society, 1*, 104-110.

Cohen, L., Verstichel, P., & Dehaene, S. (1997). Neologistic jargon sparing numbers: A category-specific phonological impairment. *Cognitive Neuropsychology, 14*, 1029-1061.

Collette, F., Majerus, S., Van der Linden, M., Dabe, P., Degueldre, C., Delfiore, G., et al. (2001). Contribution of lexico-semantic processes to verbal short-term memory tasks: A PET activation study. *Memory, 9*, 249-259.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A*, 497-505.

Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term memory recall of repeated and novel words lists. *Memory and Cognition, 21*, 539-545.

Conrad, R. (1963). Acoustic confusion and memory span for words. *Nature, 197*, 1029-1030.

Conrad, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology, 55*, 75-84.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*, 439-432.

Cowan, N. (1995). *Attention and memory: An integrated framework.* New York: Oxford University Press.

Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control (pp. 62-101).* New York: Cambridge University Press.

Cowan, N., Day, L., Saults, J. S., Kellar, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word-length on immediate memory. *Journal of Memory and Language, 31*, 1-17.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory and Cognition, 21*, 142-145.

Crutch, S. J., & Warrington, E. K. (2002). Preserved calculation skills in a case of semantic dementia. *Cortex, 38*, 389-399.

Dehaene, S. (1989). The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Perception and Psychophysics, 45*, 557-566.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, 1-42.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of numerical processing. *Mathematical Cognition, 1*, 83-120.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex, 33*, 219-250.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neuroscience, 21*, 355-361.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 626-641.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition, 43*, 1-29.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science, 284*, 970-974.

Dehaene, S., Tzourio, N., Frak, V., Raynaud, L., Cohen, L., Mehler, J., et al. (1996). Cerebral activations during number multiplication and comparison: A PET study. *Neuropsychologia, 34*, 1097-1106.

Delazer, M., & Benke, T. (1997). Arithmetic facts without meaning. *Cortex, 33*, 697-710.

Dell, G. S., & O'Seaghda, P. G. (1992). Stages of lexical access in language production. *Cognition, 42*, 287-314.

Deloche, G., & Seron, X. (1982a). From one to 1: An analysis of a transcoding process by means of neuropsychological data. *Cognition, 12*, 119-149.

Deloche, G., & Seron, X. (1982b). From three to 3: A differential analysis of skills in transcoding quantities between patients with Broca's and Wernicke's aphasia. *Brain, 105*, 719-733.

Diesfeldt, H. F. A. (1993). Progressive decline of semantic memory with preservation of number processing and calculation. *Behavioural Neurology, 6*, 239-242.

Dooling, D. J., & Lachman, R. (1971). Effects of comprehension on retention of prose. *Journal of Experimental Psychology, 88*, 216-222.

Ellis, N. C., & Hennelly, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology, 71*, 43-52.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control.* (pp. 102-134). New York: Cambridge University Press.

Epstein, W. (1961). The influence of syntactical structure on learning. *American Journal of Psychology, 74*, 80-85.

Ericsson, K. A., & Kintsch, W. (1995). Long-Term Working Memory. *Psychological Review, 102*, 211-245.

Estes, W. K. (1973). Phonemic coding and rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 12*, 360-372.

Evans, J. J., Heggs, A. J., Antoun, N., & Hodges, J. R. (1995). Progressive prosopagnosia associated with selective right temporal lobe atrophy: A new syndrome. *Brain, 118*, 1-13.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189-198.

Forde, E. M. E., & Humphreys, G. W. (2002). The role of semantic knowledge in short-term memory. *Neurocase, 8*, 13-27.

Funnell, E. (1995). Objects and properties: A study of the breakdown of semantic memory. *Memory, 3*, 497-518.

Funnell, E. (1996). Response biases in oral reading: An account of the co-occurrence of surface dyslexia and semantic dementia. *Quarterly Journal of Experimental Psychology, 49A*, 417-446.

Galton, C. J., Patterson, K., Graham, K., Lambon Ralph, M. A., Williams, G., Antoun, N., et al. (2001). Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia. *Neurology, 57*, 216-225.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition, 23*, 83-94.

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81*, 439-454.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hove, East Sussex: Lawrence Erlbaum Associates.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning Memory and Cognition, 25*, 562-562.

Gathercole, S. E., & Martin, A. J. (1996). Interactive processes in phonological memory. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 73-100). Hove, UK: Psychology Press.

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology, 54A*, 1-30.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influence of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics, 12*, 349-367.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory, 2*, 103-127.

Gevers, W., Reynvoet, B., & Fias, W. (in press). The mental representation of ordinal sequences is spatially organized. *Cognition.*

Glasspool, D. W. (1995). Competitive queuing and the articulatory loop. In P. Cairns (Ed.), *Connectionist models of memory and language* (pp. 5-30). London: UCL Press.

Graham, K. S., Hodges, J. R., & Patterson, K. (1994). The relationship between comprehension and oral reading in progressive fluent aphasia. *Neuropsychologia, 32*, 299-316.

Gregg, V. H., Freedman, C. M., & Smith, D. K. (1989). Word frequency, articulatory suppression and memory span. *British Journal of Psychology, 80*, 363-374.

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology, 44*, 339-387.

Hanten, G., & Martin, R. C. (2000). Contributions of phonological and semantic short-term memory to sentence processing: Evidence from two cases of closed head injury in children. *Journal of Memory and Language, 43*, 335-361.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language, 35*, 1-31.

Hebb, D. O. (1949). *The organisation of behavior*. New York: Wiley.

Heilman, K., Scholes, R., & Watson, R. T. (1976). Deficits of immediate memory in Broca's and conduction aphasia. *Brain and Language, 3*, 201-208.

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models in immediate serial recall. *Quarterly Journal of Experimental Psychology, 49A*, 80-115.

Hodges, J. R., Graham, N., & Patterson, K. (1995). Charting the progression in semantic dementia: Implications for the organization of semantic memory. *Memory, 3*, 463-495.

Hodges, J. R., & Patterson, K. (1995). Is semantic memory consistently impaired early in the course of Alzheimer's disease? Neuroanatomical and diagnostic implications. *Neuropsychologia, 33*, 441-459.

Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain, 115*, 1783-1806.

Howard, D., & Patterson, K. (1992). Pyramids and Palm Trees: A test of semantic access from pictures and words. Bury St. Edmunds, Suffolk: Thames Valley Test Company.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*, 685-701.

Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology, 86*, 527-536.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 1217-1232.

Huttenlocher, J., & Newcombe, N. (1976). Semantic effects on ordered recall. *Journal of Verbal Learning and Verbal Behavior, 15*, 387-399.

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior, 10*, 409-416.

Jones, G. V. (1985). Deep dyslexia, imageability and ease of prediction. *Brain and Language, 24*, 1-19.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits of long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*, 336-358.

Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic assessments of language processing in aphasia*. Hove: Lawrence Erlbaum Associates.

Knott, R. (1998). *The impact of lexical and semantic impairment on short-term verbal recall*. Unpublished thesis, University of Cambridge, Cambridge.

Knott, R., & Monsell, S. (unpublished manuscript). Words in short-term memory - what preserves their phonological structure?

Knott, R., Patterson, K., & Hodges, J. R. (1997). Lexical and semantic binding effects in short-term memory: Evidence from semantic dementia. *Cognitive Neuropsychology, 14*, 1165-1216.

Knott, R., Patterson, K., & Hodges, J. R. (2000). The role of speech production in auditory-verbal short-term memory: Evidence from progressive fluent aphasia. *Neuropsychologia, 38*, 125-142.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Lambon Ralph, M. A., Ellis, A. W., & Franklin, S. (1995). Semantic loss without surface dyslexia. *Neurocase, 1*, 363-369.

Lambon Ralph, M. A., & Howard, D. (2000). Gogi aphasia or semantic dementia? Simulating and assessing poor verbal comprehension in a case of progressive fluent aphasia. *Cognitive Neuropsychology, 17*, 437-465.

Levy, B. A. (1971). The role of articulation in auditory and visual short-term memory. *Journal of Verbal Learning and Verbal Behavior, 10*, 123-132.

Lezak, M. (1976). *Neuropsychological assessment*. New York: Oxford University Press.

Lian, A., Karlsen, P. J., & Winsvold, B. (2001). A re-evaluation of the phonological similarity effect in adults' short-term memory of words and non-words. *Memory, 9*, 281-299.

Logie, R. H., Della Sala, S., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition, 24*, 305-321.

Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language, 31*, 713-733.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*, 35-54.

Marks, L. E., & Miller, G. A. (1964). The role of semantic and syntactic constraints in the memorization of English sentences. *Journal of Verbal Learning and Verbal Behavior, 3*, 1-5.

Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. F. (1994). Origins of paraphasias in deep dyslexia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language, 47*, 609-660.

Martin, N., & Saffran, E. (1990). Repetition and verbal STM in transcortical sensory aphasia: A case study. *Brain and Language, 39*, 254-288.

Martin, N., & Saffran, E. (1992). A computational account of deep dysphasia: Evidence from a single case study. *Brain and Language, 43*, 240-274.

Martin, N., Saffran, E., & Dell, G. (1996). Recovery in deep dysphasia: Evidence for a relation between auditory-verbal STM capacity and lexical errors in repetition. *Brain and Language, 52*, 83-113.

Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology, 14*, 641-682.

Martin, R. C. (1993). Short-term memory and sentence processing: Evidence from neuropsychology. *Memory and Cognition, 21*, 176-183.

Martin, R. C., & Breedin, S. D. (1992). Dissociations between speech perception and phonological short-term memory deficits. *Cognitive Neuropsychology, 9*, 509-534.

Martin, R. C., & Lesch, M. (1996). Associations and dissociations between language impairment and list recall: Implications for models of STM. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 149-178). East Sussex, UK: Psychology Press.

Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language, 41*, 3-29.

Martin, R. C., Shelton, J., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language, 33*, 83-111.

McCarthy, R. A., & Warrington, E. K. (1987). The double dissociation of short-term memory for lists and sentences: Evidence from aphasia. *Brain, 110*, 1545-1563.

McCarthy, R. A., & Warrington, E. K. (2001). Repeating without semantics: Surface dysphasia? *Neurocase, 7*, 77-87.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114*, 159-188.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition, 44*, 107-157.

McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition, 4*, 171-196.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits of our capacity for processing information. *Psychological Review, 63*, 81-97.

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology, 63*, 176-185.

Monsell, S. (1984). Components of working memory underlying verbal skills: A "distributed capacities" view. In D. G. Bouwhuis (Ed.), *Attention and performance X* (pp. 327-350). London: Erlbaum.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature, 215*, 1519-1520.

Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S. J., & Hodges, J. R. (2000). A voxel-based morphometry study of semantic dementia: Relationship between temporal lobe atrophy and semantic memory. *Annals of Neurology, 47*, 36-45.

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology, 21*, 263-276.

Neath, I. (1997). Modality, concreteness and set-size effects in a free reconstruction of order task. *Memory and Cognition, 25*, 256-263.

Noel, M. P., & Seron, X. (1993). Arabic number reading deficit: A single case study or when 236 is read (2306) and judged superior to 1258. *Cognitive Neuropsychology, 10*, 317-339.

Noel, M. P., & Seron, X. (1997). On the existence of intermediate representations. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 697-720.

Ostergaard, A. L., & Meudell, P. R. (1984). Immediate memory span, recognition memory for subspan series of words and serial position effects in recognition memory for subspan series of verbal and nonverbal items in Broca's and Wernicke's aphasia. *Brain and Language, 22*, 1-13.

Paap, K. R., & Noel, R. W. (1991). Dual route models of print to sound: Still a good horse race. *Psychological Research, 53*, 13-24.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761-781.

Papagno, C., Valentine, T., & Baddeley, A. D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language, 30*, 331-347.

Patterson, K., Graham, N., & Hodges, J. R. (1994). The impact of semantic memory loss on phonological representations. *Journal of Cognitive Neuroscience, 6*, 57-69.

Patterson, K., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia, 30*, 1025-1040.

Patterson, K., & Hodges, J. R. (2000). Semantic dementia: One window on the structure and organisation of semantic memory. In J. Grafman (Ed.), *Handbook of neuropsychology, 2nd Edition* (Vol. 2: Memory and its disorders, pp. 313-333). Amsterdam: Elsevier.

Patterson, K., & Marcel, A. J. (1992). Phonological ALEXIA or PHONOLOGICAL alexia? In M. Radeau (Ed.), *Analytic approaches to human cognition* (pp. 259-274). Amsterdam: Elsevier.

Patterson, K., Plaut, D. C., McClelland, J. L., Seidenberg, M. S., Behrmann, M., & Hodges, J. R. (1996). Connections and disconnections: A connectionist account of surface dyslexia. In R. S. Berndt (Ed.), *Neural modelling of brain and cognitive disorders* (pp. 177-199). Singapore: World Scientific.

Peterson, L. R., & Johnson, S. T. (1971). Some effects of minimising articulation on short-term retention. *Journal of Verbal Learning and Verbal Behavior, 10*, 346-354.

Pickering, S. J., Gathercole, S. E., & Peaker, S. M. (1998). Verbal and visuospatial short-term memory in children: Evidence for common and distinct mechanisms. *Memory and Cognition, 26*, 1117-1130.

Pinel, P., Dehaene, S., Rivière, D., & LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *NeuroImage, 14*, 1013-1026.

Plaut, D., McClelland, J. L., Seidenberg, M., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes, 12*, 765-805.

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), *The emergence of language* (pp. 381-415). Mahweh, NJ: Erlbaum.

Plaut, D. C., & Shallice, T. (1991). Deep dyslexia: A case study in connectionist neuropsychology. *Cognitive Neuropsychology, 10*, 377-500.

Plaut, D. C., & Shallice, T. (1993). Preservation and semantic influences on visual object naming errors in optic aphasia: A connectionist account. *Journal of Cognitive Neuroscience, 5*, 89-117.

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology, 48A*, 384-404.

Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology, 50*, 408-412.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language, 29*, 265-282.

Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language, 38*, 265-282.

Raven, J. C. (1962). Coloured progressive matrices sets A, AB, B. London: H. K. Lewis.

Remond-Besuchet, C., Noel, M. P., Thioux, M., Brun, M., & Apse, X. (1999). Selective preservation of exceptional arithmetical knowledge in a demented patient. *Mathematical Cognition, 5*, 41-63.

Rogers, T. T., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (in press). Natural selection: The impact of semantic impairment on lexical and object decision.

Romani, C., & Martin, R. C. (1999). A deficit in the short-term retention of lexical-semantic information: Forgetting words but remembering a story. *Journal of Experimental Psychology: General, 128*, 56-77.

Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*, 29-33.

Roodenrys, S., Hulme, C., Alban, J., Ellis, A. W., & Brown, G. (1994). Effects of word frequency and age of acquisition on short-term memory span. *Memory and Cognition, 22*, 695-701.

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*, 1019-1034.

Roodenrys, S., Hulme, H., & Brown, G. (1993). The development of short-term memory span: Separable effects of speech rate and long-term memory. *Journal of Experimental Child Psychology, 56*, 431-442.

Roodenrys, S., & Quinlan, P. T. (2000). The effects of stimulus set size and word frequency on verbal serial recall. *Memory, 8*, 71-78.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General, 126*, 211-227.

Rossor, M. N., Warrington, E. K., & Cipolotti, L. (1995). The isolation of calculation skills. *Journal of Neurology, 242*, 78-81.

Ruchkin, D. S., Berndt, R. S., Johnson, R., Jr., Grafman, J., Ritter, W., & Canoune, H. L. (1999). Lexical contributions to retention of verbal information in working memory: Event-related brain potential evidence. *Journal of Memory and Language, 41*, 345-364.

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (in press). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences.*

Rummer, R., & Engelkamp, J. (2001). Phonological information contributes to short-term recall of auditorily presented sentences. *Journal of Memory and Language, 45*, 451-467.

Rummer, R., & Engelkamp, J. (2003). Phonological information in immediate and delayed sentence recall. *The Quarterly Journal of Experimental Psychology, 56A*, 83-95.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics, 2*, 437-442.

Saffran, E. M., & Martin, N. (1990). Neuropsychological evidence for lexical involvement in short-term memory. In T. Shallice (Ed.), *Neuropsychological impairments of short-term memory* (pp. 428-447). Cambridge, UK: Cambridge University Press.

Saint-Aubin, J., & Poirier, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *Quarterly Journal of Experimental Psychology, 52A*, 367-394.

Saint-Aubin, J., & Poirier, M. (2000). Immediate serial recall of words and nonwords: Tests of the retrieval based hypothesis. *Psychonomic Bulletin & Review, 7*, 332-340.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory and Cognition, 21*, 168-175.

Schweickert, R., Chen, S., & Poirier, M. (1999). Redintegration and the useful lifetime of the verbal memory representation. *International Journal of Psychology, 34*, 447-453.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *The Quarterly Journal of Experimental Psychology, 51A*, 283-304.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

Shallice, T., & Butterworth, B. (1977). Short-term memory impairment and spontaneous speech. *Neuropsychologia, 15*, 729-735.

Shallice, T., & Vallar, G. (1990). The impairment of auditory-verbal short-term storage. In T. Shallice (Ed.), *Neuropsychological impairments of short-term memory* (pp. 11-53). Cambridge, UK: Cambridge University Press.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology, 22*, 261-273.

Shulman, H. G. (1971). Similarity effects in short-term memory. *Psychonomic Bulletin, 75*, 399-415.

Snowden, J. S., Goulding, P. J., & Neary, D. (1989). Semantic dementia: A form of circumscribed cerebral atrophy. *Behavioural Neurology, 2*, 167-182.

Snowden, J. S., Griffiths, H., & Neary, D. (1994). Semantic dementia: Autobiographical contribution to preservation of meaning. *Cognitive Neuropsychology, 11*, 265-288.

Snowden, J. S., Griffiths, H. L., & Neary, D. (1996). Semantic-episodic memory interactions in semantic dementia: Implications for retrograde memory function. *Cognitive Neuropsychology, 13*, 1101-1137.

Snowden, J. S., & Neary, D. (2002). Relearning of verbal labels in semantic dementia. *Neuropsychologia, 40*, 1715-1728.

Snowden, J. S., Neary, D., & Mann, D. M. A. (Eds.). (1996). *Frontotemporal lobar degeneration: Frontotemporal dementia, progressive aphasia, semantic dementia.* London: Churchill Livingstone.

Stanescu-Cosson, R., Pinel, P., van de Moortele, P., Le Bihan, D., Cohen, L., & Dehaene, S. (2000). Understanding dissociations in dyscalculia: A brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain, 123*, 2240-2255.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 1140-1154.

Takarama, Y., Sugishita, M., Akiguchi, I., & Kimura, J. (1994). Isolated acalculia due to left parietal lesion. *Archives of Neurology, 51*, 286-291.

Tehan, G., & Humphreys, M. S. (1988). Articulatory loop explanations of memory span and pronunciation rate correspondences: A cautionary note. *Bulletin of the Psychonomic Society, 26*, 293-296.

Thioux, M., Pillon, A., Samson, D., de Partz, M. P., Noel, M. P., & Seron, X. (1998). The isolation of numerals at the semantic level. *Neurocase, 4*, 371-389.

Thorn, A. S. C., & Gathercole, S. E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *The Quarterly Journal of Experimental Psychology, 52A*, 303-324.

Thorn, A. S. C., Gathercole, S. E., & Frankish, C. R. (2002). Language familiarity effects in short-term memory: The role of output delay and long-term knowledge. *The Quarterly Journal of Experimental Psychology, 55A*, 1363-1383.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 145-152.

Trojano, L., & Grossi, D. (1995). Phonological and lexical coding in verbal short-term memory and learning. *Brain and Language, 51*, 336-354.

Trojano, L., Stanzione, M., & Grossi, D. (1992). Short-term memory and verbal learning with auditory phonological coding deficit: A neuropsychological case study. *Brain and Cognition, 18*, 12-33.

Turner, J. E., Henry, L. A., & Smith, P. T. (2000). The development of the use of long-term knowledge to assist short-term recall. *The Quarterly Journal of Experimental Psychology, 53A*, 457-478.

Tyler, L. K., Voice, J. K., & Moss, H. E. (1996). The interaction of semantic and phonological processing. *Proceedings of the Eighteen Annual Conference of the Cognitive Science Society.*

Vallar, G., & Baddeley, A. D. (1984). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior, 23*, 151-161.

Vallar, G., & Papagno, C. (2002). Neuropsychological impairments of verbal short-term memory. In B. A. Wilson (Ed.), *The Handbook of Memory Disorders.* Oxford: John Wiley & Sons, Ltd.

Van Bon, W. H. J., & Van der Pijl, J. M. L. (1997). Effects of word length and wordlikeness on pseudoword repetition by poor and normal readers. *Applied Psycholinguistics, 18*, 101-114.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound and reading. *Memory and Cognition, 15*, 181-198.

Venezky, R. L. (1970). *The structure of English orthography.* The Hague, The Netherlands: Mouton.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 1256-1271.

Warrington, E. K. (1975). The selective impairment of semantic memory. *The Quarterly Journal of Experimental Psychology, 27*, 635-657.

Warrington, E. K. (1982). The fractionation of arithmetical skills: A single case study. *Quarterly Journal of Experimental Psychology, 34A*, 31-51.

Warrington, E. K., & James, M. (1991). The Visual Object and Space Perception battery. Bury St. Edmunds, Suffolk: Thames Valley Test Company.

Warrington, E. K., & Shallice, T. (1969). The selective impairment of auditory-verbal short-term memory. *Brain, 92*, 885-896.

Watkins, M. J. (1977). The intricacy of memory span. *Memory and Cognition, 5*, 529-534.

Watkins, O. C., & Watkins, M. J. (1977). Serial recall and the modality effect: Effects of word frequency. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 712-718.

Wechsler, D. (1987). Wechsler Memory Scale - Revised (WMS-R). New York: Psychological Corporation.

Wechsler, D. (1997). Wechsler Memory Scale - Third Edition (WMS-III). London: The Psychological Corporation.

Wetherick, N. E. (1975). The role of semantic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 471-480.

Willis, C. S., & Gathercole, S. E. (2001). Phonological short-term memory contributions to sentence processing in young children. *Memory, 9*, 349-363.

Wright, C. C. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory and Cognition, 7*, 411-419.

List of abbreviations

ANOVA: Analysis of variance

ISR: Immediate serial recall

LTM: Long-term memory

PDP: Parallel distributed processing

SD: Semantic dementia

STM: Short-term memory

WM: Working memory

Appendix 1: Average properties of the known and degraded words selected for each patient in Experiment 1, Chapter 2

	Set size	Lemma frequency (Celex)	Frequency (Kucera & Francis, 1967)	Imageability	Syllable length
EK Known	11	18.6	15.7*	590*	2.2
EK Degraded		19.1	15.2*	591	1.7
GT Known	9	18.0	14.6*	560*	2.0
GT Degraded		18.3	14.0*	584*	2.0
PD Known	6	59.9	28.4	587	1.6
PD Degraded		51.7	27.8	614	1.6
MK Known	7	76.1	54.1	587*	1.1
MK Degraded		85.1	52.0	602	1.1

* denotes that values were unavailable for some items

Appendix 2: Average properties of the known and degraded words selected for each patient in Experiment 2, Chapter 2

	Set size	Frequency (Celex)	Frequency (Kucera & Francis, 1967)	Imageability	Syllable length
EK Known	15	39.3	30.5	419	2.3
EK Degraded		37.8	38.6	390	2.4
GT Known	14	54.9	48.5	434	2.1
GT Degraded		53.3	50.7	405	2.2
PD Known	14	43.6	43.1	480	1.9
PD Degraded		44.6	37.7	447	2.2

Appendix 3: Average properties of the known and degraded words selected for each patient in Experiment 3, Chapter 2

	Set size	Frequency (Celex)	Frequency (Kucera & Francis, 1967)	Imageability	Syllable length
EK Known	23	95.7	64.8*	587*	1.7
EK Degraded		107.7	30.0*	592*	1.7
GT Known	22	138.5	113.1*	584*	1.6
GT Degraded		20.7	14.1*	591*	1.6
MK Known	26	169.3	168.9	599*	1.3
MK Degraded		24.2	33.2*	589*	1.3

* denotes that values were unavailable for some items

Appendix 4: Average properties of the known and degraded words selected for each patient in Experiment 5, Chapter 2

	Frequency (Celex)	Frequency (Kucera & Francis, 1967)	Imageability	Syllable length
EK Known	23.5	22.9*	559	2.1
EK Degraded	23.9	23.5*	492	2.1
GT Known	34.8	36.1*	508*	1.8
GT Degraded	32.1	31.6*	487*	1.8

* denotes that values were unavailable for some items

Appendix 5: The two word sets matched to single-digit numbers in Experiment 1, Chapter 3

	Digits	Frequency-matched words	Frequency and imageability-matched words
	One	All	Her
	Two	Well	Back
	Three	Though	Small
	Four	Lot	Name
	Five	Soon	Light
	Six	Road	Age
	Seven	Sorry	Council
	Eight	Worth	Health
	Nine	Bread	Sight
Mean syllable length	1.1	1.1	1.1
Mean frequency	748.9	786.8	799.7
Mean imageability	449.8	439.2*	458.0

* score is unavailable for some items

Appendix 6: Low frequency multi-digit numbers and matched words used in Experiment 2, Chapter 3

	Numbers	Frequency-matched	Frequency-matched, high imageability
	Eleven	Article	Furniture
	Thirteen	Birthday	Tennis
	Seventeen	Definite	Envelope
	Nineteen	Trading	Dusty
	Seventy	Notably	Pollution
	Eighty	Shiny	Cigar
	Ninety	Applause	Cement
	Billion	Gesture	Novel
	Trillion	Blandly	Madman
Mean syllable length	2.3	2.3	2.3
Mean frequency	23.4	26.5	24.3
Mean imageability	Not available	Not available	564.4

Appendix 7: Number and face-part words used in Experiment 3, Chapter 3

	Numbers	Frequency-matched words
	Eight	Mouth
	Nine	Neck
	Ten	Hair
	Eleven	Tongue
	Twelve	Nose
	Thirteen	Brow
	Fourteen	Beard
	Sixteen	Chin
	Eighteen	Forehead
	Nineteen	Fringe
	Sixty	Cheek
	Seventy	Parting
Mean syllable length	1.8	1.25
Mean frequency	61.1	62.5
Mean imageability	Not available	599.7*

* score is unavailable for some items

Appendix 8: The long and short words used to examine the effect of word length in Chapter 4

Short words: guy, break, tall, dress, jump, league, desk, dust, aid, hard, buy, scene, score, bomb, dream, dance, beach, arm, pair, moon, rock, stay, tree, wait, close, sea, rain, blue, care, form, taste, cause, brain, trade, lock, beat, truck, roll, nose, add.

Long words: confidence, poetry, appearance, reference, instruction, excellent, newspaper, realize, consider, century, liberal, existence, typical, foundation, location, condition, assignment, discussion, properly, orchestra, telephone, decision, sensitive, construction, position, popular, universe, physical, medical, national, establish, soviet, encourage, addition, particle, electric, completion, arrangement, minister, assistance

Appendix 9: The characteristics of the known and degraded words used in Chapter 4

	Frequency (Celex)	Frequency (Kucera & Francis, 1967)	Imageability	Mono-syllabic words†	Two syllable words†	Words with three+ syllables†
SJ Known	27.3	27.0*	540*	12	14	6
SJ Degraded	28.1	20.0*	514*			
EK Known	43.7	31.9*	593*	13	15	8
EK Degraded	42.9	38.6*	510*			
KI Known	23.8	19.4*	596*	13	15	8
KI Degraded	21.3	17.0*	597*			
JT Known	71.3	45.4*	595*	16	2	2
JT Degraded	67.3	43.1	596*			
GT Known	54.3	42.9*	556*	10	12	6
GT Degraded	54.0	38.4*	534*			

* denotes that values were unavailable for some items

† Number of items in each set of known or degraded words

Appendix 10: The word and nonword stimuli used in Chapter 5, Experiment 1

Words: High frequency, high imageability

ball, bed, board, boat, book, face, fish, foot, girl, gun, head, hill, home, horse, house, king, leg, male, men, neck, night, park, phone, rain, rock, seat, teeth, wall, wife, wine

Words: High frequency, low imageability

base, date, death, form, god, half, hope, jack, job, lead, line, loss, mass, name, part, race, role, rule, shape, sharp, shock, side, size, term, thing, thought, till, top, voice, week

Words: Low frequency, high imageability

bat, boot, cane, cart, coin, dime, dove, fan, foam, fog, geese, harp, hawk, hedge, hen, hoof, jeep, kite, limb, mouse, noose, rat, rib, surf, thorn, toad, web, weed, wig, yacht

Words: Low frequency, low imageability

bang, bet, bid, curse, dip, foul, germ, hurt, hush, jade, kale, knoll, lean, lodge, mall, mash, psalm, rack, raid, rhyme, sage, sap, thud, ton, verb, vice, wharf, whiff, wrath, zone

Nonwords

baf, bal, bam, barl, barss, beel, beng, beuffe, bick, bim, bol, bon, boof, bot, bowne, burge, berl, cowt, cun, dap, deef, dem, dibe, dit, dop, dorth, fak, fal, feem, feen, fet, fid, fik, fing, foate, fod, fok, fon, forp, garl, gen, gid, girse, gis, goyt, haid, hal, han, harg, heem, heen, hees, heff, hess, het, hin, hoak, hoat, hoess, hol, hom, hon, houne, hoys, hud, hus, jarm, jid, jong, jook, jote, jud, jurn, jurz, kang, keem, keet, kep, kerm, kerze, ket, koese, korp, lan, laysh, leet, lep, lidge, lif, loate, lood, looth, lut, mal, med, mek, min, moess, morke, mort, mot, mun, nate, ned, neek, noid, nood, nooke, nop, paim, pid, poeth, rab, rad, raig, raim, raish, raowl, rel, roak, roarss, rork, rorl, rorm, rorn, rorsch, rud, ruuge, sawg, saybe, sayde, saysh, seipe, seithe, sek, sharf, sharl, shart, sherb, sherp, siebe, siefe, sisle, sorl, taybe, tayde, tayje, tayne, tayse, tayve, tharj, tharss, thayte, thert, thit, thoape, tice, tiefe, tiege, vayze, vike, vipe, vite, voan, wais, wann, warthe, weem, weis, werp, wid, woam, woan, woash, wol, wole, wote, yourss, zime, zine

Appendix 11: Additional stimuli used in Chapter 5, Experiment 2

Words: High frequency, high imageability

book, cash, dark, dog, feet, gun, heart, king, love, mouth, nose, park, pool, red, road, roof, room, ship, shop, song, sun, teeth, wheel, white, wood

Words: High frequency, low imageability

call, cut, feel, fell, fine, firm, lead, long, lord, mean, miss, move, part, piece, rise, save, south, thick, turn, type, wait, warm, wide, wish, work,

Words: Low frequency, high imageability

cage, cart, cave, chalk, cheese, duck, geese, gem, gym, heel, juice, lamb, leaf, limb, morgue, nail, noose, peach, pearl, pet, pig, shed, thumb, wig, wool

Words: Low frequency, low imageability

bait, bang, cheat, chic, curse, dirge, hail, jerk, kale, keel, latch, lodge, loon, meek, nerve, nip, pawn, push, rap, rung, sod, tuck, verb, whack, whoop

Note: Some of these items also appear in Appendix A. The mixed lists that contained a single word were not tested as pure word lists, allowing their items to be reused.

Nonwords

bav, bayth, baz, bem, besh, buthe, bip, boash, boove, borch, borf, dayss, dess, dieje, dif, dutt, fam, fape, feuke, fiss, foade, foage, fodge, forsh, fout, geeth, girfe, goz, gudge, harn, harss, heg, hep, herch, heth, hidge, hoad, hobe, horg, horp, howke, huth, huthe, jad, jaim, jeese, jerss, jod, karch, kaych, kaysh, kerg, kidge, kieze, koite, leeb, leeth, lesh, liepe, ling, lom, lov, mav, mave, maz, mep, mord, mout, naze, ness, nide, nooth, noz, pard, parsh, rabe, radge, raing, ral, rayfe, riesh, rizz, roaje, rol, rosh, ruebe, saf, sape, sarg, seeve, sieje, sert, sharb, shayse, shoss, sime, teep, thib, thorb, thorm, thush, tiss, tope, tov, tud, turp, verp, vime, voig, weck, weeb, weef, weige, widge, wime, wiv, worg, worg, yod, zope

Appendix 12: Additional stimuli used in Chapter 5, Experiment 3

High frequency, high imageability words

bag, bath, coat, court, fall, farm, food, hall, hell, light, meal, moon, note, page, sign, team, town, van, walk, youth

High frequency, low imageability words

cause, course, deal, good, guess, keep, lack, look, loss, main, make, might, need, nice, rate, right, role, sight, while, wrong

Low frequency, high imageability words

barn, bud, dive, fork, goat, hog, jet, keg, lace, moth, pill, pine, pit, pole, ram, rug, shawl, sword, tomb, wreck

Low frequency, low imageability words

booth, cope, dell, fail, fill, foal, hide, join, lathe, lease, loan, myth, pat, peel, rush, sop, thong, toil, toll, zeal

Appendix 13: Matching span stimuli used with healthy participants in Chapter 6, section 6.3

High frequency, high imageability words

Changes between two words

phone	dus	**bag**	hame	sherl
light	hal	feep	**bed**	sim
lunn	**well**	jong	**board**	hoas
shan	**book**	deeth	mel	**court**
buv	nid	**shop**	sart	**hell**
male	dipe	**seat**	kess	bVsh
rock	beej	gid	**sun**	lurm
fuv	**ball**	raig	**team**	shid
pake	**fish**	rUl	korp	**white**
bup	deeve	**hill**	nart	**wood**

Changes between a word and a nonword

gun	kal	DOG	**bife**	torm
sign	fum	**lorg**	CASE	hoat
feese	**wife**	bon	HEAD	**noyl**
lorf	**nose**	mep	**shol**	FOOT
HORSE	bipe	**men**	naze	**poil**
home	fot	MOUTH	**kun**	werg
town	hoaf	**dup**	WALL	vate
holl	**fat**	mun	LEG	**koess**
lort	**youth**	kidge	**tQn**	RED
ROOM	mon	**note**	gade	**biff**

Items that exchange: **bold type**

Words, unchanging: UPPER CASE

Nonwords, unchanging: lower case

Celex code: V-ʌ, U-ʊ, Q-ɒ

Changes between two nonwords

RAIN	**theed**	LOVE	**joll**	bime
WALK	**rowl**	lem	HEART	**deesh**
waythe	MINE	**nUl**	FEET	**tass**
lorp	NIGHT	**mang**	turl	FARM
werje	shate	HALL	**feem**	NECK
FACE	**thunn**	TEETH	**weff**	rorl
FALL	**liff**	kaig	WINE	**hoosh**
thoat	FOOD	**wess**	KING	**parl**
ket	ROAD	**feeg**	guyle	HOUSE
fith	hoak	GIRL	**jaowed**	PAGE

High frequency, low imageability words

Changes between two words

lack	poam	**bit**	feeve	woole
need	woam	pooss	**fine**	rarl
lork	**deal**	fis	**rate**	mide
kerm	**side**	mot	beeth	**role**
haid	Tan	**week**	fol	**size**
look	j6d	**turn**	shart	waim
make	fiD	leet	**while**	pon
soat	**term**	yun	**wide**	ral
taid	**sight**	werve	mek	**hope**
rorg	b6th	**name**	lat	**god**

Changes between a word and a nonword

thing	deeje	MIGHT	**soam**	fup
thick	fUsh	**pell**	CAUSE	beess
geed	**long**	thail	FORM	**sighje**
nurl	**keep**	sade	**dudge**	WRONG
PIECE	vade	**half**	meg	**lorm**
lead	sorm	PART	**raybe**	futh
short	pid	**taig**	LOSS	han
roass	**till**	fom	LORD	**kem**
lon	**base**	feek	**pem**	RIGHT
FORCE	geel	**shape**	roak	**tith**

Items that exchange: **bold type**

Words, unchanging: UPPER CASE

Nonwords, unchanging: lower case

Celex code: 6-aʊ, T-θ, U-ʊ, D-ð

Changes between two nonwords

CALL	**poat**	MISS	**feeth**	werb
TALK	**fooss**	lart	JOB	**wal**
rorf	MASS	**haig**	FEEL	**bine**
widge	MAIN	**sime**	bol	COURSE
murl	hon	SAFE	**rorb**	NICE
WARM	**feethe**	SOUTH	**hik**	lurt
VOICE	**rark**	gort	MOVE	**feege**
harl	TYPE	**jock**	RACE	**beeve**
libe	MEAN	**toag**	sorl	FIRM
hooss	sem	THOUGHT	**tiv**	WORK

Low frequency, high imageability words

Changes between two words

dive	kooss	**hen**	vert	mork
fork	moess	vid	**cart**	paim
fack	**goat**	thorl	**cage**	hes
mort	**hoof**	kile	boip	**ram**
kime	lek	**hose**	jid	**pearl**
leaf	hoad	**thorn**	thaze	kell
noose	rol	dibe	**jet**	yaid
shet	**pole**	tiss	**hawk**	wan
p2g	**rib**	het	torss	**nail**
s21	larp	**wig**	thaid	**pet**

Changes between a word and a nonword

bat	hime	COIN	**nork**	roaf
foam	rorss	**nurt**	HEDGE	tal
shoat	**cane**	vel	SWORD	**lipe**
lowss	**bud**	jaowt	**mip**	WRECK
MOUSE	thate	**limb**	feesh	**joff**
keg	shaowed	LOOT	**porm**	fon
dime	bod	**laig**	FAN	hUl
nade	**pill**	fUss	HARP	**moag**
f2the	**rat**	sork	**von**	DOVE
WHIP	fush	**geese**	yan	**nake**

Items that exchange: **bold type**

Words, unchanging: UPPER CASE

Nonwords, unchanging: lower case

Celex code: U-ʊ, 2-aI,V-ʌ

Changes between two nonwords

BEAN	**hong**	SHAWL	**waipe**	fidge
MOTH	**reet**	waig	PINE	**gell**
miv	FOG	**sal**	BARN	**dife**
waithe	BOOT	**sike**	junn	TOAD
werp	boose	SHED	**kaot**	RUG
LACE	**thoad**	TOMB	**weck**	deethe
LARK	**foosh**	woess	SURF	**maig**
tid	WEB	**rorm**	HOG	**buyve**
kaiz	PIT	**fung**	hool	LAMB
raim	fVl	WEED	**thert**	MORGUE

Low frequency, low imageability words

Changes between two words

kale	yat	**dip**	sorf	beesh
bang	loat	deek	**foul**	wid
woab	**psalm**	torp	**keel**	rife
keet	**mash**	worss	jomm	**loan**
searle	weess	**bard**	gann	**lodge**
ton	loak	**peel**	hurm	wowed
toil	paiz	fide	**rack**	bQth
kade	**zone**	liss	**sap**	fime
pode	**lean**	h6t	sayze	**whoop**
heek	norl	**sage**	fub	**win**

Changes between a word and a nonword

bid	wote	LATHE	**mal**	sonn
tuck	joth	**sarl**	COPE	bUs
lan	**fill**	rUs	HURT	**nike**
voad	**curse**	meb	**tordge**	THONG
BEAT	hurk	**save**	foad	**nool**
loon	murt	FAIL	**baoge**	sark
lease	boosh	**seff**	GERM	rart
rooss	**dell**	fiv	HIDE	**sorp**
heg	**jade**	bot	**rork**	ZEAL
SOD	wole	**bet**	fip	**d2the**

Items that exchange: **bold type**

Words, unchanging: UPPER CASE

Nonwords, unchanging: lower case

Celex code: U-ʊ, 2-aɪ, V-ʌ, 6-aʊ, Q-ɒ

Changes between two nonwords

thit	DIRGE	**jopp**	kang	**thit**
dutt	gack	MYTH	**lep**	**dutt**
RHYME	**leb**	TOLL	**hidge**	RHYME
NIP	**torg**	kide	VERB	NIP
hort	RUSH	**wem**	VICE	hort
toess	WHACK	**bozz**	hile	**toess**
nide	bVl	WHARF	**jowt**	**nide**
JOIN	**bVth**	WHIFF	**serm**	JOIN
HUSH	**fipe**	lorss	WRATH	HUSH
roosh	SOP	**deef**	THUD	roosh